ISSN: 2083-389X
eISSN: 2084-3127

© J Hear Sci, 2022; 12(4): 55–67
DOI: 10.17430/JHS.2022.12.4.7

INDO-SPRITT: DEVELOPMENT AND NORMS OF AN INDONESIAN SPEECH RECOGNITION THRESHOLD TEST FOR CHILDREN

Contributions:
A Study design/planning
B Data collection/entry
C Data analysis/statistics
D Data interpretation
E Preparation of manuscript
F Literature analysis/search
G Funds collection

Dahlia Sartika[1,2A-G], Philip Newall[1,2AC-G], Harvey Dillon[1,3AC-F]

[1] Department of Linguistics, Macquarie University, Australia
[2] NextSense Institute, Australian Hearing Hub, Macquarie University, Australia
[3] Manchester Centre for Audiology and Deafness, University of Manchester, United Kingdom

Corresponding author: Dahlia Sartika, Department of Linguistics, Macquarie University, Macquarie Park NSW 2109, Australia; email: dahlia.sartika@nextsense.org.au

Abstract

Introduction: Well-designed audiometric speech tests for Indonesian children are not currently available. This paper describes the development of the Indonesian Speech Recognition Threshold Test (INDO-SPRITT).

Material and methods: A list of Indonesian words with response foils and pictures was developed. Presentation level was varied and the 50% recognition threshold was calculated as the average of the midpoints of each reversal. A normative reference was established using a sample of 118 normal hearing participants,16 children with severe to profound hearing loss, and 25 adults. The effects of age on speech reception thresholds and test reliability were also assessed.

Results: INDO-SPRITT material was found to be appropriate for children older than 4 years and 6 months. The speech reception threshold (SRT) improved on average from 18 dB HL for 4 to 5 year old children to 13 dB for children aged 10 to 13 years, providing a normative reference against which the SRT of children with unknown hearing status can be compared. Five reversals are enough to estimate the SRT.

Conclusions: Suitable words, phonemic balance, and pictures have been created for Indonesian children. The reliability of different lengths of the test was similar, with 5 reversals being enough to estimate the SRT. The mean SRT decreased with age, but did not vary with the number of reversals.

Key words: speech perception • SRT • children • Indonesian language

INDO-SPRITT: OPRACOWANIE I NORMY INDONEZYJSKIEGO TESTU PROGOWEGO ROZPOZNAWANIA MOWY DLA DZIECI

Streszczenie

Wprowadzenie: Obecnie nie ma dobrze zaprojektowanego audiometrycznego testu słownego dla dzieci indonezyjskich. Ta praca opisuje proces opracowywania Indonezyjskiego Testu Progowego Rozpoznawania Mowy (Indonesian Speech Recognition Threshold Test, INDO-SPRITT).

Materiał i metody: Opracowano listę słów indonezyjskich z kartami odpowiedzi i ilustracjami. Poziom prezentacji materiału był zróżnicowany, próg rozpoznawania na poziomie 50% został obliczony jako średnia punktów środkowych każdego powtórzenia. Referencyjne wartości normatywne zostały ustalone na podstawie wyników badań na próbie 118 uczestników z normalnym słuchem, 16 dzieci z niedosłuchem ciężkim do głębokiego oraz 25 dorosłych. Oceniono także wpływ wieku na progi percepcji mowy i rzetelność testu.

Wyniki: Wyniki pokazują, że INDO-SPRITT jest odpowiedni dla dzieci w wieku powyżej 4 lat i 6 miesięcy. Próg percepcji mowy (SRT) poprawił się średnio o 18 dB HL w grupie dzieci w wieku 4–5 lat i o 13 dB HL w grupie dzieci w wieku 10–13 lat. Wartości te mogą być wykorzystane jako wartości referencyjne, do których można porównywać wyniki SRT dzieci o nieznanym stanie słuchu. Pięć powtórzeń wystarcza do oceny SRT.

Wnioski: Dla indonezyjskich dzieci opracowano odpowiedni zestaw słów, zrównoważonych fonematycznie i z ilustracjami. Rzetelność różnych długości testu była podobna, pięć powtórzeń wystarcza do oceny SRT. Średnia wartość SRT zmniejszała się z wiekiem, ale nie zmieniała się zależnie od ilości powtórzeń.

Słowa kluczowe: percepcja mowy • SRT • dzieci • język indonezyjski

Introduction

There is a need to develop standardised speech audiometry materials for children in the Indonesian language (Bahasa Indonesia, BI), although there are some non-standardised pediatric speech test materials for screening school age children. Word lists are presented by means of live voice at whispered voice level in an open set condition and the children are asked to repeat the word heard [1]. However, this test format is not suitable for use with younger children. Furthermore, an oral response is inappropriate for those children with articulation problems, because the tester cannot be sure whether a response is due to faulty hearing or faulty articulation or both [2]. Therefore, our aim was to develop standardised speech test material for children using a more appropriate response mode.

In the United States, the preferred materials for measurement of speech reception threshold (SRT) are spondaic words. In theory, however, almost any material can be used [3]. Most of the words in BI are disyllabic or polysyllabic, and BI does not have spondaic or trochaic words because stress is essentially free and communicatively irrelevant.

Although there are more than 700 ethnic languages in Indonesia [4], in this study the national language, Bahasa Indonesia (BI), is used. BI is the language of government and the medium of instruction in schools and is used in an increasingly wide sphere of social interaction, including inter-ethnic communication, religion, and mass communication. There is an increasingly large population of speakers for whom BI is their first language. An estimated 23 million people speak BI as a first language and an additional 140 million speak it as a second [5,6]. Most children in big cities speak BI for daily communication. However, many children in the villages or small towns usually speak their ethnic language for daily communication. Therefore, it is intended that the Indonesian Speech Reception Threshold Test (INDO-SPRITT) described in this paper would mainly be used in big cities such as Jakarta, Surabaya, Yogyakarta, and other cities where children use BI for daily communication.

The phonological basis of BI comprises 22 consonant phonemes, six monophthong vowel phonemes, and three diphthongs, /ai/, /oi/, /au/ [7,8].

Some principles in the development of speech audiometry for children are different than they are for adults: (1) cognitive, motoric, and attentional demands of the test should be age-appropriate; (2) the task must be interesting and motivating; (3) the test should place minimal demands on vocabulary knowledge and higher-level language ability; (4) the test should not require phonological knowledge or speech production skills [9-11].

Using pictures to represent the test items and their foils facilitates testing of very young children. In a picture pointing response, the child simply points to one of several pictures associated with the word they heard, instead of repeating or writing it. This approach minimizes the effect of any articulation difficulties and places no demand on writing ability. It is important that the picture accurately represents each test item [12,13].

INDO-SPRITT was developed using Northwestern University Children's Perception of Speech (NU-CHIPS) [14] as the basis of its development. NU-CHIPS was developed as a simple test which involves presenting recorded word lists to a child and asking the child to point to a picture of the word they heard. The target picture is presented with three foils that have some auditory resemblance to the target.

The speech test described in this study determines the speech level at which the test participant correctly recognises 50% of the presented speech material. This is referred to as the Speech Recognition Threshold (SRT).

This paper establishes normative data for the INDO-SPRITT, against which the results for children with unknown hearing ability can be compared [15,16] and examines its reliability as the test length is varied.

Methods

Development of INDO-SPRITT material

Developing word lists

INDO-SPRITT was designed using the basic principles for speech tests but taking into account the special characteristics of BI. Because a well-established resource of word counts in BI for children was not available, the words used in INDO-SPRITT were chosen from 29 Indonesian children's books. The books consisted of story books with pictures intended for children in kindergarten and first year primary school.

From these 29 books, 6837 words were extracted. The Tact software [17] was used to determine word frequency and phoneme frequency. Exactly 266 two-syllable concrete words (those for which a picture could be used to represent the target word or response foils, e.g., *anjing*/dog) were selected (**Appendix A**).

Then 50 of the most frequent concrete words were selected to make up word lists for the test material. The selection was based on the following criteria: (1) Words with a high frequency of occurrence were used; ideally the first 50 high-frequency words should be used, but in this study this principle could not always be applied because it was not possible to fulfill all the other criteria (see below). (2) The selected word had to have at least one similarly sounding word that also occurred with a high frequency. If any word was unsuitable, the word with the next lowest level frequency was chosen. (3) To find one or more similarly sounding words, word structures that were as similar as possible were used (see later). (4) The 50 words must be phonemically balanced. This meant the different phonemes should occur in the test materials in approximately the same frequency as the ones in the sample language from the children's books.

As with NU-CHIPS [14], another three-word lists were developed using the same 50 words in a different random order in each list (**Appendix B**). Thus, INDO-SPRITT has four different test forms. This method of using the same words in every list with only the order changed is one approach to achieving list equivalence [18]. For meaningful stimuli, this approach seems suitable only when the speech test is to be used to determine the speech recognition threshold (SRT), rather than the maximum attainable intelligibility (because significant learning of the stimuli by the participant is inevitable when they are presented with high intelligibility).

Developing the foils

The test items are inserted among foil items that sound similar to the test item. Such response alternatives are referred to as effective foils [19]. The foils were selected from the 266-word corpus of concrete words. Foils were chosen based on words with high frequency, phonetic similarity, similarity in word structure to the test items, and representable by pictures.

An example of a test item with foils that met all of the above criteria is *sapi* (cow) and its foils *nasi* (rice), *dasi* (tie), and *bayi* (baby). These four words have a high familiarity (based on their frequency), phonetic similarity in terms of vowel sound and word structure (CV-CV), and are representable by pictures.

However, finding foils with word structure similar to the test items in Indonesian bisyllabic words is not always easy. For example, for the test item *bola* (ball), the foils we chose are *domba* (lamb), *tomat* (tomato), and *coklat* (chocolate). In this example, the foils have phonetic similarity in terms of vowel sound, but the word structures are different from the test item. The word structure of *bola* is CV-CV, *domba* is CVC-CV, *tomat* is CV-CVC, and *coklat* is CV-CCVC. Additionally, in bisyllabic words there are many more possible consonants than vowels. The pattern of consonants in Indonesian words is complex because of both the number of consonants and the number of places where consonants can occur.

In all cases, the foils have the same vowel sound as the test item. Also, the foils for each item are the same in each list.

Developing the pictures

The pictures of the 50 target words and their foils were drawn by an artist based on those used in Indonesian children's books. At the initial stage of the study, the picture books were in black and white in order to be consistent with the picture books in NU-CHIPS. The pictures were then assessed by the participants and then modified if they were not age-appropriate.

Field studies

The field studies consisted of two parts: Part 1 to assess the familiarity of the INDO-SPRITT material, and Part 2 to determine normative data and to assess reliability as test length was varied.

Participants

57 normally hearing children, 16 children with severe to profound hearing loss, and 10 normally hearing adults participated in Part 1.

Table 1. Participants by age and number

Age group	Number
Two-year-olds	11
Three-year-olds	25
Four-year-olds	11
Five-year-olds	4
Six-year-olds	2
Seven-year-olds	4

Of the 57 normally hearing children, 24 were assessed in a hospital setting. They had hearing thresholds better than 20 dB HL in at least one ear, at each frequency from 500 to 4000 Hz, and with type A tympanograms. Their ages ranged from 2 to 7 years. The other 33 children were assessed in two preschools. Their hearing was presumed to be normal, or close to normal, on the basis of a teacher's report that they had normal speech and language development. A hearing test was not performed in the preschools because the teachers were unfamiliar with hearing tests and were concerned that the equipment might harm the children. The teachers only allowed the author to perform the test with the INDO-SPRITT picture books and no other equipment. Although this was far from ideal, it is reasonable to assume that the children's hearing was normal or close to normal in at least one ear.

The number of participants within each age groups depended on the children available at the time of testing and was variable (**Table 1**). The child participants were tested in Indonesia while adult participants were tested in Sydney, Australia.

The 16 children with severe to profound hearing loss all wore hearing aids and were recruited from a school for the deaf of 80 students in Jogjakarta. Their ages ranged from 4 to 12 years. Recent audiograms of the children were obtained from the school principal.

The 10 normally hearing adults were postgraduate students from our university. All were native Indonesian speakers and used the language for daily communication. They had hearing thresholds better than 20 dB HL in at least one ear, at each frequency from 250 to 8000 Hz with type A tympanograms. These adults participated in the evaluation of picture familiarity (both in the original and modified versions).

In Part 2 of the study there were 61 normally hearing children (aged 4.5 to 13 years) and 15 normally hearing adults. (The children who were tested in the first phase did not participate in the second). They had hearing threshold level no worse than 20 dB HL in at least one ear, at each frequency from 500 to 4000 Hz, with type A tympanograms. The children lived in or near Jakarta and all used BI for their daily communication. The 15 normally hearing adults were selected as in Part 1. The purpose of using adults was to confirm that the test worked before testing it on children.

Equipment

Speech test materials described earlier were used. The audiometric assessments used an Oscilla 960 audiometer, TDH 39 headphones, a GSI 36 tympanometer, a CD player attached to the audiometer, and a sound level meter.

Part 1

Part 1 involved assessment of the familiarity of the INDO-SPRITT material

Word assessment (test items and its foils)

The assessment was conducted in a relatively quiet room in all the places mentioned above. A sound level meter was used to ensure that the ambient noise did not exceed 50 dB(A). The child sat about 1 m in front of the tester and the words (both targets and foils) were presented by live voice. The child was asked to listen carefully to the word and to look at all the pictures on the page and to point to the picture corresponding to the word that was heard. After the child pointed to the picture, the tester turned the page and asked the child to listen to the next word. The child was encouraged to make a guess even if he/she was not sure of the response. The words were repeated if the child did not respond within a few seconds.

After testing the INDO-SPRITT material on normal hearing participants, and after appropriate amendments were made to the pictures, the modified materials (test items and foils) were assessed on the hearing impaired participants. Initially, assessment was planned to be done on participants with different degrees of hearing loss, but the majority of children in the institutions for hearing impaired children had severe to profound sensorineural loss.

Initially the procedures for the hearing-impaired participants were similar to those used with the normal hearing participants. However, none of the participants could perform the task (familiarity assessment score was close to zero). As a consequence, the participants were permitted to lip-read and most of the time a written version of the word also had to be presented (and the familiarity assessment score improved). This was done because the intention of this part of the study was to assess the children's familiarity with the words and not to assess their sensory ability.

After the completion of testing the familiarity of the INDO-SPRITT materials, the modified word lists (spoken by the first author during familiarity testing) were recorded using the first author's voice. Pictures of the 50 words and pictures of the foils were redrawn in colour by an artist.

Picture assessment (test item and its foils)

When a word was not familiar to a child, or when a child did not point to any picture or pointed to the wrong picture, it was necessary to investigate whether it was the word or the picture that the child did not recognise. For example, if a child did not point to any picture when a word of *hidung* (nose) was administered, the child was then asked to point to her own nose. If the child could point to their nose, it meant that the picture was the problem, not the word. The pictures that were not recognised or unfamiliar to the children were modified. The pictures (original and modified) were then assessed with 10 adult participants.

Part 2

Part 2 involved the collection of normative and reliability data.

Pre-test

Prior to the SRT test, play audiometry and tympanometry were undertaken with each child. After this, the words with the modified pictures were shown to each child to ensure that the modified pictures were easily recognised. All participants had hearing thresholds no worse than 20 dB HL and normal tympanograms.

SRT test

The child was instructed to listen through the earphones to a CD player of a woman saying some words to the child. Then the child was asked to listen carefully to the word and to look at all four pictures on the page and to point to the picture corresponding to the word that was heard. After the child pointed to a picture, the tester turned the page and asked the child to listen to the next word. The child was instructed to listen carefully since some of the words might be very soft, and they were encouraged to make a guess if they were not sure of the answer.

A list of 50 words was tested in each ear using an adaptive procedure [20]. At the initial stage of the study the starting point of the SRT test was 30 dB SL (relative to 3FAHL – three-frequency average of thresholds at 500, 1000, and 2000 Hz) since Elliot & Katz [14] found ceiling effects at this level. However, it was found that for a few children 30 dB SL was not loud enough to get all responses correct. Thus, it was decided to start the SRT test at 40 dB SL for the rest of this study.

When a child gave a correct response, the following stimulus was decreased by 5 dB until they gave an incorrect response (the first reversal). The presentation level was then increased by 5 dB until the child gave a correct response (the second reversal). At least 15 reversals (i.e., upward or downward) were performed. The better ear was tested first, then the test was continued using the other ear. If the child was tired or not able to continue, the test was ended and therefore only one ear was tested. A different test form was randomly chosen for each child and each ear.

A comparison of the speech threshold for different numbers of reversals (5, 10, and 15) was performed. The standard error of measurement (based on midpoints of 5, 10, and 15 reversals) and width of tracking excursion between each pair of reversals were calculated. For each pair of reversals, the excursion width was equal to the dB difference between the upper and lower limits of that excursion. The test procedure for the adult participants was the same as for the children, except that the adults were asked to repeat the words that they heard (an open set test).

Table 2. List of 50 words ranked according to their frequency (the total number of times each word appeared in the books). Words changed after preliminary testing are shown with an asterisk

Frequency	Words	Frequency	Word
49	Rumah (house)	10	Topi (hat)
44	Ikan (fish)	10	Lilin (candle)
36	Kaki (foot)	10	Kursi (chair)
27	Tangan (hand)	10	Coklat (chocolate)
27	Buku (book)	10	Nanas (pineapple)
26	Kucing (Cat)	9	Gajah (elephant)
26	Bayi (baby)	9	Guling (body pillow)
25	Pisang (banana)	9	Tikus (mouse)
25	Bola (ball)	9	Piring (plate)
24	Burung (bird)	8	Dasi (tie)
23	Anak (child)	8	Anjing (dog)
21	Payung (umbrella)	7	Bantal (pillow)
21	Bunga (flower)	7	Garpu (fork)
18	Mata (eye)	7	Lampu (lamp)
17	Buah (fruit)	7	Bapak (father)
16	Bulan (moon)	6	Gayung (scoop)
16	Ayam (chicken)	6	Ember (bucket)
13	Sapi (cow)	6	Semut (ant)
13	Jari (finger)	6	Sandal (sandals)
11	Nasi (rice)	5	Gunting (scissors)
11	Badut (clown)	5	Kambing (goat)
11	Rambut (hair)	5	Kapal (ship)*
11	Roti (bread)	4	Bebek (duck)
11	Hidung (nose)	4	Cecak (gecko)*
11	Gelas (glass)	4	Telur (egg)

Results

Part 1: Assessment of familiarity with the INDO-SPRITT material

Normally hearing participants

Word assessment

Two words, kera (monkey) and katak (frog), were highly unfamiliar to many 2 to 4-year-old children. The children were more familiar with their alternative words. Kodok was an alternative word for katak and monyet for kera. However, those words could not be used because they did not sound similar to the target word. Therefore, they were replaced by cecak (gecko) and kapal (ship). These words were the next lower level frequency of words in the 266 word corpus of concrete words.

A return visit was made to the school and the replacement words were tried with 19 of the 3-year-old children; all of them were able to recognise the new words. As a result of changing the two test words (*kapal* and *cecak*, see **Table 2**), the frequency of some phonemes such as /k/, /t/, /r/, /l/, /p/, and /c/ changed (**Table 3**).

Foil assessment

The ideal number of effective foils is three for each test item. However, the outcome of the familiarity assessment resulted in a situation where some test items had less than three effective foils (i.e., they did not share the same vowels as the target word, as is the case with NU-CHIPS). This was because some foils were not familiar to the children and due to the unavailability of concrete words that met the criteria mentioned in the methods section. Therefore, there was the following variation in the number of effective foils: (i) two test items with one effective foil; (ii) six

Table 3. Comparison between phoneme frequency in the base material (left) and phoneme frequency in the test material after the assessment study (right)

Phoneme	Phoneme frequency in the base material (%)	Phoneme frequency in the test material (%)
/a/	19.1	19.3
/i/	18.1	9.7
/u/	6.9	9.7
/ə/	6.1	1.8
/k/	6.1	5.7
/n/	5.1	4.8
/t/	5.1	3.1
/r/	4.8	4.4
/m/	4.6	3.5
/l/	3.8	4.4
/b/	3.8	6.1
/s/	3.7	4.4
/ŋ/	3.4	5.7
/d/	3.2	1.7
/p/	3.0	3.1
/h/	2.6	1.3
/o/	1.8	1.8
/g/	1.4	2.6
/e/	1.3	1.8
/j/	1.2	0.9
/y/	0.9	0.9
/c/	0.8	1.8

test items with two effective foils; and (iii) 42 test items with three effective foils.

Picture assessment

For many of the participants several original pictures were not recognised or were confused with other alternative pictures. As an example, some children pointed to the picture of an orange (**Figure 1A,** indicated by red arrow) when the word 'telur' (egg) was heard. According to parents and teachers, the children are more likely to recognise an egg if it is an oval shape, rather than a broken one. So, the egg picture was then replaced with a whole egg oval shape in the context of a chicken (**Figure 1B**).

In summary, we found that (1) 2-year-olds had problems with seven pictures: telur (egg), mata (eye), hidung (nose), jari (finger), sandal (sandal), guling (body pillow) and kambing (goat). (2) 3-year-olds had problems with five pictures: telur, jari, hidung, anjing and kambing. (3) 4-year-olds had problems with two pictures: hidung and rambut (hair). (4) 5-year-olds had problems with one picture: telur. (5) 6- and 7-year-olds had no problems.

The result of picture assessments with 10 adult participants was that all of them agreed with the name of the pictures and were able to recognize all pictures. This indicates that the problems occurred because some of the pictures were not age-appropriate.

Table 4 shows that up to and including the age of 4 years, 50% or more of the children demonstrated difficulty in identifying a picture out of context or an abstract representation of the object, whereas from 6 years onwards these difficulties were no longer present.

Hearing-impaired participants

Table 5 shows the percentage of modified pictures that were recognized by each of the 16 severely to profoundly hearing-impaired children. The pictures were accurately recognized by all the children 7 years of age or older once they knew the word. This contrasted with the children

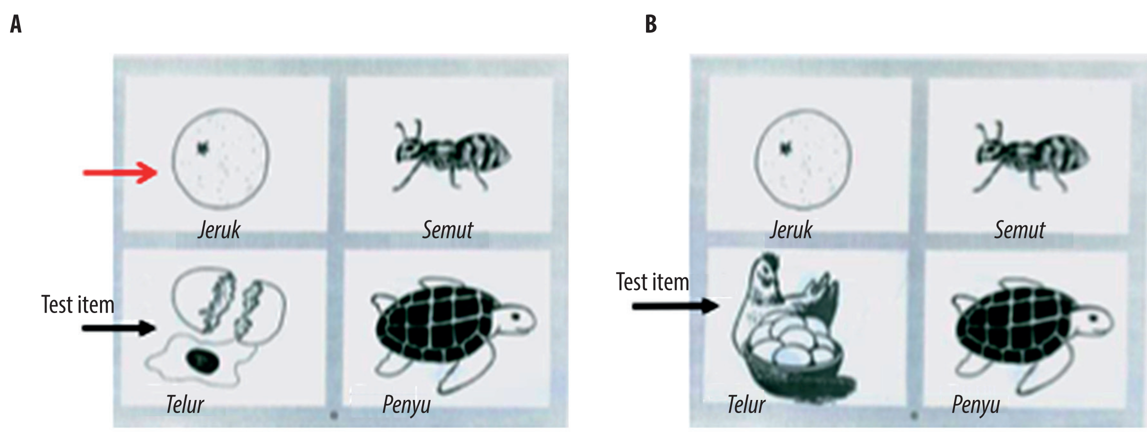

Figure 1. Initial illustration (**A**) of *telur* (egg), and its improved replacement (**B**)

Table 4. The number of participants of each age (n) and the number of those who were unable to recognize one or more of the original pictures (n_p)

Age (years)	n	n_p	%
2	11	11	100
3	25	13	52
4	11	3	27
5	4	1	25
6	2	0	0
7	4	0	0
Adult	0	0	0

Table 5. Familiarity scores (%) of the hearing-impaired children on the modified material

Subject	Age (years, months)	Familiarity score (%)
S1	4, 11	60
S2	5, 8	76
S3	5, 10	88
S4	6, 5	90
S5	7, 6	98
S6	7, 6	98
S7	8, 3	100
S8	8, 4	100
S9	8, 10	100
S10	9	100
S11	9	100
S12	9, 4	100
S13	10	100
S14	10, 2	98
S15	10, 7	98
S16	12, 2	100

with normal hearing who were able to recognize the pictures at a younger age.

Part 2: Normative study and reliability assessment

Normative study

Figure 2 shows the mean SRT for each age group for SRTs based on either 5, 10, or 15 reversals. The mean SRT decreases as age increases, but it did not vary with the number of reversals.

For the adults who were presented the sounds without being shown the pictures, the mean SRT also did not vary with the number of reversals.

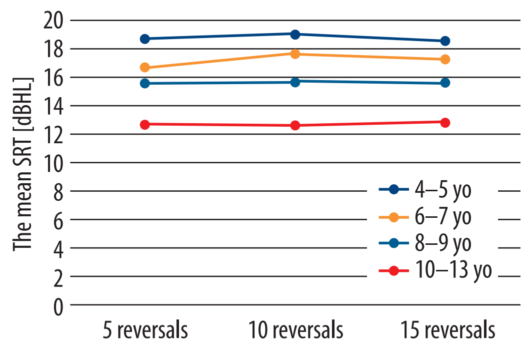

Figure 2. SRT for the children across four age groups based on 5, 10, or 15 reversals

The effect of age and number of reversals on the SRT was examined with a two-way ANOVA with age as a between-groups independent variable, and number of reversals as a repeated-measures independent variable. The ANOVA showed that the main effect of age was significant ($F(3,228) = 13.2$; $p < 0.001$) but that the number of reversals was not ($F(2, 228) = 0.03$; $p = 0.97$).

Reliability and test length

Standard error of measurement

Standard error (SE) values were calculated for each participant's threshold after 5, 10, or 15 reversals by dividing the standard deviation of the reversal midpoints by the square root of the number of midpoints. The resulting standard errors, averaged across participants, are presented in **Figure 3**.

The SE measurements decreased as the number of reversals increased, as expected. The mean SE for 5 reversals was 1.2, for 10 reversals it was 1.0, and for 15 reversals it was 0.9 dB.

Width of the tracking excursion

Finally, the width of the tracking excursion was considered as an index of test reliability. For each participant, these excursions were averaged across reversals to give a mean excursion for each participant.

The average excursion width for children was 8.0 dB (with standard deviation 4.8 dB) and for adults it was also 8.0 dB (with standard deviation 4.9 dB). Given the step-size of 5 dB used in this study, the results indicate that typically each participant required an increase or decrease in level for around two words before the next reversal occurred. **Figure 4** shows that after the first few excursions (i.e., reversals), the excursion width did not vary as the test progressed.

Discussion

This paper has described the development and normalization of INDO-SPRITT, a speech identification test for children in Bahasa Indonesia which uses a closed response set of pictures. It has four alternative forms and uses an

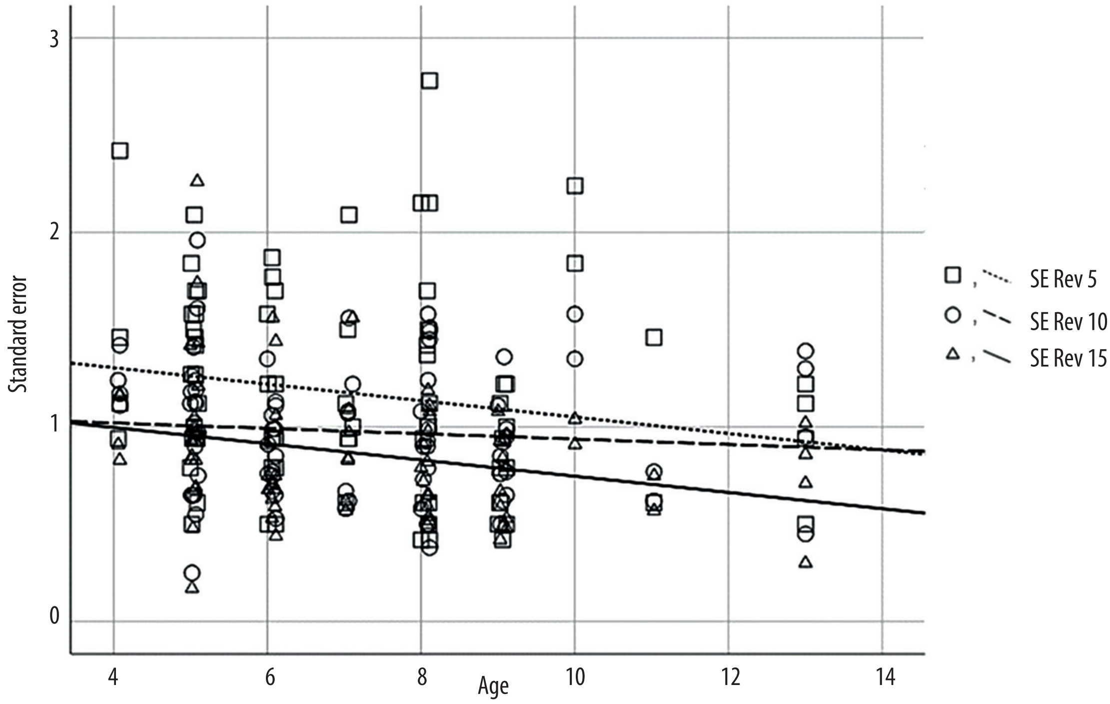

Figure 3. Standard error measurement for the children, as a function of age, for different numbers of reversals

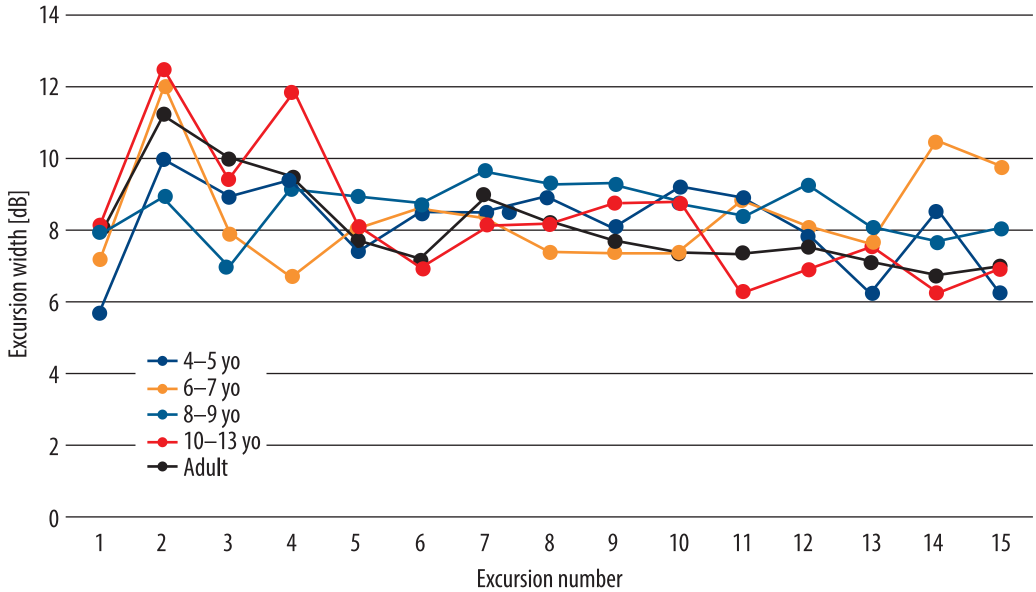

Figure 4. Mean excursion width on INDO-SPRITT for each age group over 15 reversals

approach in which each list has the same words with only the order changed. INDO-SPRITT has been designed to limit the effect of receptive language ability on test performance, as well as the effect of extra-auditory factors such as physical (articulation) problems and educational (writing) ability.

With the sub-set of profoundly hearing-impaired children a wide range of cues (auditory, lip reading, and written) was required for assessing the familiarity of the INDO-SPRITT test items. Otherwise, none of the words could be recognised by the participants. This might have occurred because INDO-SPRITT is a closed-set speech recognition test with response foils that have similar syllabic structure and share some phonemes. The experiments indicated that profoundly deaf children, who can only hear vowel sounds, or can only perceive the suprasegmental aspects, could not reliably be tested with this material as they could not tell the difference between the words.

The test item is inserted among foil items representing selected phonemic confusions, but wherever possible, using the same vowels as the target. The ideal number of effective foils is three for each test item; this number is important, because the probability of selecting a foil depends on the number of possible responses [10,21]. For example, with one foil the chance score is 50%, for two foils it is 33%, and so on. Therefore, with fewer foils the scores are likely to be better.

Even though there was some variation in the number of effective foils for each test item, most of the test items in the INDO-SPRITT have three effective foils, and this variation is the same for each list, unlike the situation with NU-CHIPS [14] where the number for an item can vary from list to list. The matching words in NU-CHIPS are based on inconsistent criteria in terms of test item and foil; in some cases, consistency between the test item and the foil lies just in the vowel, but in others it is just the last or first consonant, and in others several sounds are consistent.

For children under the age of 4 years, 50% or more of the children had difficulty in identifying a picture out of context or an abstract representation of the object. This observation can be explained in two ways [22]. The first is context. The younger the child the more likely it is that they have difficulty identifying a picture out of context or which is abstract. They are also more influenced by how a word is presented diagrammatically and the context in which the word has been learned. The second factor is concept learning. Children over 6 years old have enough life experience to be able to apply knowledge across various situations – for example, recognizing an eye on the face in a drawing as a single element. However, in children under the age of 4 years, 50% or more will have difficulties. All these factors will impact upon the ability of the child to recognize a picture and to match a word to it. These issues were solved by modifying the pictures based on the assessments.

Although little is known or discussed in the audiology literature about the importance of how pictures are perceived in the construction of speech tests for young children, this study showed that it is important to consider such effects in constructing picture-based speech tests for this population.

Although a long test can be expected to produce the best test precision, using only 5 reversals for test administration is an advantage because a fast clinical method is usually more popular with audiologists or other hearing health care professionals than a long method, due to limited times available for testing in a clinic. Furthermore, a fast and efficient measurement is important in a child population due to their limited attention span. With only five reversals, the standard error of measurement was most commonly around 1 dB, which is not much different compared to the standard error for 10 and 15 reversals. Furthermore, two-way ANOVA showed that the effect of the number of reversals on SRT was not significant ($F(2,228) = 0.030$; $p = 0.97$). Thus, 5 reversals are enough to estimate SRT.

The SRTs measured with INDO-SPRITT are able to provide an objective and quantitative measure of a communication disability [17]. After medical treatment or fitting of a hearing aid, SRTs can be re-measured to give an indication of the extent to which the disability has been reduced.

INDO-SPRITT materials consist of the same words and foils in every list with only the order changed. Because of this, possible learning effects may be of concern. However, when this type of material is used to determine SRT, learning the stimuli by the participant is unlikely [18]. In its intended use, the intensity is always at a level where the words are just recognizable, and therefore it is harder to learn the stimuli.

Elliot and Katz [14] showed that learning did not occur when the same words in every list, with different randomization, were presented to a listener in the same test session. This learning effect study was carried out using presentation levels of 0 and 2 dB SL relative to normal hearing children's SRT.

There may be a concern that pictures used in this study may contribute to the learning effect. However, when the children point to a picture associated with the word heard, they are not told whether they are right or wrong, hence their memory is not reinforced by feedback. So, even though they may see a picture several times, it is unlikely that memory will significantly contribute to the learning effect.

Two-way ANOVA did show that there was a significant age effect. The mean SRT decreased as age increased, a result consistent with other studies. Elliot and Katz (1980) found that at sensation levels lower than 30 dB (0 and 2 dB SL), an age effect was evident, so that 10-year-olds performed better than 5-year-olds who in turn performed better than 3-year-olds. The likely explanation for the age effect observed at low sensation levels is language skill or experience. Mackie and Dermody [20] found that children could recognize all words when they were presented at a normal conversational level, but when these same words were presented at low speech levels, as in the level used for SRT, 3-year-olds showed a higher threshold than 5-year-olds, and 5-year-olds showed a higher threshold than 7-year-olds. Elliot et al. [23] found that developmental improvement in performance could be explained by "word frequency effects", which means that children knew and were familiar with the stimulus words even though they had less experience with them compared to adults. Due to these frequency effects, young children (under the age of 8) needed speech to be at a higher intensity level than for older children or adults in order for the word to be understood [24].

Random errors of measurement are never completely eliminated, and the degree of error in the test consequently needs to be known. A measurement can be labelled as reliable when the amount of random error is slight [25]. The calculated standard error for 5 reversals was only 1.2 dB. This is likely to be a slight underestimate of the true standard error for repeated administration of the test because the mid-points of adjoining reversals are unlikely to be statistically independent of each other.

The use of speech and picture representations that we have shown are highly familiar to children contribute to the face, content, and construct validity of the test. It is realised that male and female voices may influence a child's speech perception. Female voices are usually softer and occupy higher frequency ranges than male voices [26] with the latter about an octave lower in pitch than the former [27]. In the presence of high-level noise, the intelligibility of female voices is less than of male voices [28], but under ordinary noise and listening conditions the intelligibility of male and female speech is approximately equal. The purpose of the test is to differentiate children with different degrees of hearing loss, and as this can be done with voices of either gender, we do not consider that the use of just a female voice limits the usefulness of the test.

Conclusions

The experiments allow us to conclude that:

- INDO-SPRITT is appropriate for children older than 4 years and 6 months, although its testing procedures should not be used with severely to profoundly hearing-impaired participants who can only hear vowel sounds or who can only perceive suprasegmental aspects.
- The normative reference value of SRT ranges from 7 to 26 dB HL, and 5 reversals are enough to estimate SRT with an estimated standard error of around ± 1 dB.
- The words, phonemic balance, and pictures appear suitable for Indonesian children.
- It is important to consider both developmental word recognition and picture effects when constructing a speech test for this population.

Acknowledgments

We highly appreciate the substantial contribution by Dr Robert Mannell and Dr Mridula Sharma at Macquarie University. This work was supported by the Rotary Club of Carlingford, NSW, Australia. We also appreciate assistance by Robyn Cantle Moore of the NextSense Institute and Peter Bosloper at Otometrics Australia. Harvey Dillon acknowledges the support of Macquarie University and the NIHR Manchester Biomedical Research Centre.

Declaration of interest

The authors report no conflicts of interest. The authors alone are responsible for the content and writing of the paper.

Appendix A
List of 266 concrete words based on frequency order

49 rumah	9 batu	5 gunting	3 dahan	2 lantai	1 hewan
44 ikan	9 daun	5 gunung	3 danau	2 lidah	1 injil
33 kera	9 gajah	5 ibu 3	duyung	2 monyet	1 intan
36 kaki	9 guling	5 jagung	3 itik	2 motor	1 kendi
27 buku	9 orang	5 jahe	3 jeruk	2 nangka	1 kertas
27 tangan	9 perut	5 kambing	3 kakek	2 obat	1 keset
26 bayi	9 piring	5 kapal	3 keong	2 otak	1 ketel
26 kucing	9 tikus	5 kerbau	3 lalat	2 padi	1 kijang
25 bola	8 anjing	5 kompor	3 leher	2 paha	1 kodok
25 pisang	8 bintang	5 kuku	3 lutut	2 paku	1 koran
24 burung	8 dasi	5 muka	3 naga	2 panah	1 korek
24 loli	8 mangga	5 obeng	3 nenek	2 pipa	1 kunci
23 anak	8 sendok	5 panci	3 penyu	2 poci	1 laso
21 bunga	8 wajan	5 pisau	3 permen	2 pusar	1 lembu
21 payung	7 air	5 ubi	3 punggung	2 quran	1 limau
20 pensil	7 api	5 yoyo	3 rumput	2 ranting	1 lonceng
20 pohon	7 bantal	4 bajing	3 sapu	2 robot	1 manggis
20 baju	7 garpu	4 bebek	3 sawah	2 roda	1 mangkok
18 mata	7 kayu	4 cecak	3 sayur	2 rusa	1 martil
17 buah	7 lampu	4 foto	3 tumit	2 sayap	1 mawar
17 gigi	7 langit	4 handuk	3 ulat	2 serbet	1 merak
17 mulut	7 nyamuk	4 jambu	3 wayang	2 tangga	1 mesjid

16 ayam	7 pintu	4 kado	2 atap	2 tiang	1 panda
14 kancil	7 wortel	4 obor	2 bakso	2 unta	1 pena
13 jari	7 tali	4 pagar	2 bangau	2 vandel	1 pesut
13 meja	7 bapak	4 palu	2 cangkul	2 weker	1 pompa
13 sapi	6 ember	4 pantai	2 capung	1 anggur	1 rantai
13 tubuh	6 gayung	4 raja	2 cincin	1 angklung	1 ratu
12 tupai	6 lebah	4 salak	2 dadu	1 asbak	1 roket
11 badut	6 musang	4 sangkar	2 dahi	1 awan	1 rokok
11 gelas	6 odol	4 sarang	2 gitar	1 cacing	1 rubah
11 hidung	6 pantat	4 siku	2 golok	1 candi	1 sawo
11 mobil	6 papan	4 sisir	2 guci	1 ceret	1 sekop
11 nasi	6 sabun	4 susu	2 hiu	1 cobek	1 senter
11 roti	6 sandal	4 telur	2 jamur	1 congklak	1 singa
11 rambut	6 semut	4 udang	2 jarum	1 cula	1 siput
10 ayah	6 sikat	4 zebra	2 kail	1 delman	1 sumur
10 coklat	6 tulang	3 babi	2 kapur	1 domba	1 surat
10 katak	6 ular	3 bambu	2 kasur	1 dompet	1 tomat
10 kursi	6 tulang	3 batang	2 kaus	1 dongkrak	1 tugu
10 lilin	5 biji	3 benang	2 kompor	1 kuku	1 wajah
10 nanas	5 cabai	3 bibir	2 koper	1 fliper	
10 topi	5 dada	3 botol	2 kuda	1 gasing	
10 badan	5 ekor	3 cermin	2 kumbang	1 gaun	
9 apel	5 elang	3 daging	2 labu	1 gua	

Appendix B
Four test forms

	Susunan Tes A1		*Susunan Tes A2*		*Susunan Tes B3*		*Susunan Tes B4*
	Test Form A1		***Test form A2***		***Test form B3***		***Test form B4***
1	Bantal (pillow)	1	Mata (Eye)	1	Lilin (Candle)	1	Lilin (Candle)
2	Kaki (Foot)	2	Bayi (Baby)	2	Topi (Hat)	2	Roti (Bread)
3	Bulan (Moon)	3	Bunga (Flower)	3	Kursi (Chair)	3	Gunting (Scissors)
4	Mata (Eye)	4	Bantal (Pillow)	4	Badut (Clown)	4	Garpu (Fork)
5	Ayam (Chicken)	5	Katak (Frog)	5	Roti (Bread)	5	Topi (Hat)
6	Badut (Clown)	6	Rambut (Hair)	6	Nasi (Rice)	6	Anjing (Dog)
7	Sandal (Slippers)	7	Nanas (Pineapple)	7	Ember (Bucket)	7	Bebek (Duck)
8	Gajah (Elephant)	8	Gajah (Elephant)	8	Bola (Ball)	8	Coklat (Chocolate)
9	Piring (Plate)	9	Lilin (Candle)	9	Anak (Child)	9	Katak (Frog)
10	Buku (Book)	10	Burung (Bird)	10	Bantal (Pillow)	10	Nanas (Pineapple)
11	Nanas (Pineapple)	11	Sandal (Slippers)	11	Rambut (Hair)	11	Badut (Clown)
12	Roti (Bread)	12	Roti (Bread)	12	Ayam (Chicken)	12	Anak (Child)

	Susunan Tes A1		*Susunan Tes A2*		*Susunan Tes B3*		*Susunan Tes B4*
	Test Form A1		***Test form A2***		***Test form B3***		***Test form B4***
13	Semut (Ant)	13	Semut (Ant)	13	Nanas (Pineapple)	13	Mata (Eye)
14	Kera (Monkey)	14	Gelas (Glass)	14	Ikan (Fish)	14	Pisang (Banana)
15	Ember (Bucket)	15	Bebek (Duck)	15	Gelas (Glass)	15	Gelas (Glass)
16	Pisang (Banana)	16	Pisang (Banana)	16	Gajah (Elephant)	16	Tangan (Hand)
17	Jari (Finger)	17	Jari (Finger)	17	Dasi (Tie)	17	Kaki (Foot)
18	Bayi (Baby)	18	Kaki (Foot)	18	Rumah (House)	18	Rumah (House)
19	Buah (Fruit)	19	Buah (Fruit)	19	Bulan (Moon)	19	Bunga (Flower)
20	Bapak (Father)	20	Anak (Child)	20	Lampu (Lamp)	20	Lampu (Lamp)
21	Dasi (Tie)	21	Dasi (Tie)	21	Guling (–)	21	Kucing (Cat)
22	Bunga (Flower)	22	Rumah (House)	22	Telur (Egg)	22	Semut (Ant)
23	Anak (Child)	23	Ayam (Chicken)	23	Bapak (Father)	23	Ayam (Chicken)
24	Katak (Frog)	24	Bapak (Father)	24	Bebek (Duck)	24	Ember (Bucket)
25	Rambut (Hair)	25	Garpu (Fork)	25	Sapi (Cow)	25	Sapi (Cow)
26	Payung (Umbrella)	26	Gayung (Bath scoop)	26	Bayi (Baby)	26	Bayi (Baby)
27	Lilin (Candle)	27	Piring (Plate)	27	Pisang (Banan)	27	Ikan (Fish)
28	Burung (Bird)	28	Buku (Book)	28	Piring (Plate)	28	Piring (Plate)
29	Topi (Hat)	29	Topi (Hat)	29	Payung (umbrella)	29	Gayung (Bath scoop)
30	Gelas (Glass)	30	Kera (Monkey)	30	Bunga (Flower)	30	Buah (Fruit)
31	Telur (Egg)	31	Telur (Egg)	31	Tikus (Mice)	31	Hidung (Nose)
32	Bebek (Duck)	32	Ember (Bucket)	32	Mata (Eye)	32	Bantal (Pillow)
33	Ikan (Fish)	33	Ikan (Fish)	33	Tangan (Hand)	33	Gajah (Elephant)
34	Kursi (Chair)	34	Kursi (Chair)	34	Gunting (Scissors)	34	Kursi (Chair)
35	Anjing (Dog)	35	Kambing (Goat)	35	Coklat (Chocolate)	35	Bola (Ball)
36	Lampu (Lamp)	36	Badut (Clown)	36	Buku (Book)	36	Buku (Book)
37	Tangan (Hand)	37	Tangan (Hand)	37	Anjing (Dog)	37	Kambing (Goat)
38	Guling (–)	38	Gunting (Scissors)	38	Buah (Fruit)	38	Bulan (Moon)
39	Sapi (Cow)	39	Anjing (Dog)	39	Burung (Bird)	39	Burung (Bird)
40	Gayung (Bath scoop)	40	Payung (Umbrella)	40	Kera (Monkey)	40	Kera (Monkey)
41	Garpu (Fork)	41	Lampu (Lamp)	41	Garpu (Fork)	41	Rambut (Hair)
42	Kucing (Cat)	42	Guling (–)	42	Kucing (Cat)	42	Guling (–)
43	Coklat (Chocolate)	43	Bola (Ball)	43	Katak (Frog)	43	Bapak (Father)
44	Tikus (Mice)	44	Hidung (Nose)	44	Gayung (Bath scoop)	44	Payung (Umbrella)
45	Bola (Ball)	45	Coklat (Chocolate)	45	Jari (Finger)	45	Dasi (Tie)
46	Kambing (Goat)	46	Nasi (Rice)	46	Kaki (Foot)	46	Jari (Finger)
47	Hidung (Nose)	47	Tikus (Mice)	47	Semut (Ant)	47	Telur (Egg)
48	Gunting (Scissors)	48	Kucing (Cat)	48	Hidung (Nose)	48	Tikus (Mice)
49	Rumah (House)	49	Bulan (Moon)	49	Sandal (Slippers)	49	Sandal (Slippers)
50	Nasi (Rice)	50	Sapi (Cow)	50	Kambing (Goat)	50	Nasi (Rice)

References

1. Soewito A, Djoko SS, Soejarno AH. Pembakuan Tes Bisik Kata bisilabik Bahasa. Yogyakarta: Universitas Gajah Mada; 1984.
2. Markides A. Speech tests of hearing for children. In: Martin M, editor. Speech Audiometry. London: Taylor & Francis; 1987, 155–70.
3. Stach BA. Speech Audiometry. In: Stach BA, editor. Clinical Audiology: An Introduction. San Diego: Singular;1998, 229–52.
4. Lewis MP. Ethnologue: Languages of the World. Dallas: SIL International; 2009.
5. Gordon RG. Ethnologue: Languages of the World. Dallas: SIL International; 2005.
6. Grimes CE. Indonesian: the official language of a multilingual nation. In: Wurm SA, Muhlhausler P, Tyron DT, editors. Atlas of Languages of Intercultural Communication in the Pacific, Asia, and the Americas (Trends in Linguistics. Documentation 13). Berlin: Mouton de Gruyter; 1996, 719–27.
7. Lapoliwa H. A generative approach to the phonology of Bahasa Indonesia. Canberra: Department of Linguistics, Research School of Pacific Studies, Australian National University; 1981.
8. Soderberg CD, Olson KS. Indonesian. Journal of the International Phonetic Association: Illustrations of the IPA, 2008; 38(2): 209–13.
9. Jerger S. Speech Audiometry. In: Jerger J, editor. Pediatric Audiology. San Diego: College Hill Press; 1984.
10. Kirk KI, Diendorf AO, Pisoni DB, Robbins AM. Assessing speech perception in children. In: Mendel LL, Danhauer JL, editors. Audiologic Evaluation and Management and Speech Perception Assessment. San Diego: Singular; 1997, 101–31.
11. Mendel LL, Pousson MA, Bass JK, Coffelt JA, Morris M, Lane KA. Spanish Picture Identification Test. Am J Audiol, 2020; 29(3): 318–28.
12. Kendall DC. Audiometry for young children. The Teacher of the Deaf, 1954; LII/52(307), 18–23.
13. Mendel LL, Pousson MA, Bass JK, Lunsford RE, McNiece C. Spanish Pediatric Speech Recognition Threshold (SPSRT) Test. Am J Audiol, 2019; 28(3): 597–604.
14. Elliot LL, Katz DR. Development of a New Children's Test of Speech Discrimination. St Louis: Auditec; 1980.
15. Doyne MP, Steer MD. Studies in speech reception testing. J Speech Hear Disord, 1951; 16(2): 132–9.
16. Soli SD. Some thoughts on communication handicap and hearing impairment. Int J Audiol, 2008; 47(6): 285–6.
17. Bradley J, Lancashire I, Presutti L, Stairs M. Text Analysis Computing Tools (TACT), version 2.1.4; 1995.
18. Dillon H, Ching T. What makes a good speech test? In: Plant G, Spens KE, editors. Profound Deafness and Speech Communication. London: Whurr Publishers; 1995.
19. Black JW. Responses to multiple-choice intelligibility tests. J Speech Hear Res, 1968; 11: 453–66.
20. Mackie K, Dermody P. Use of a monosyllabic adaptive speech test (MAST) with young children. J Speech Hear Res, 1986; 29(2): 275–81.
21. Mackie KC, Dermody PJ. Word intelligibility tests in audiology for the assessment of communication adequacy. Canberra: National Acoustic Laboratories (Australia); 1982.
22. Wechsler D. The Wechsler Preschool and Primary Scale of Intelligence (WPPSI-III). San Antonio: The Psychological Corporation; 2002.
23. Elliot LL, Clifton LA, Servi DG. Word frequency effects for a closed-set word identification task. Audiology, 1983; 22(3): 229–40.
24. Byrne D. Word familiarity in speech perception testing of children. Australian J Audiol, 1983; 5(2): 77–80.
25. Mendel LL, Danhauer JL. Test development and standardization. In: Mendel LL, Danhauer JL, editors. Speech Perception Assessment. San Diego: Singular; 1997, 7–15.
26. Fletcher H. Speech and Hearing. New York: Van Nostrand; 1929.
27. French NR, Steinberg JC. Factors governing the intelligibility of speech sounds. J Acoust Soc Am, 1947; 19: 90–119.
28. Silverstein B, Bilger RC, Hanley TD, Steer MD. The relative intelligibility of male and female talkers. J Educ Psychol, 1953; 44(7): 418–28.