Development and Standardization of Arabic Version of Bamford-Kowal- Bench Speech in Noise Test for Children

Original Article

Ghada Gamal Amer, Enaas Ahmed Kolkaila, Trandil Hassan Elmahallawy, Amani Mohamed El-Gharib

Department of Audio-vestibular Medicine, Faculty of Medicine, Tanta University, Egypt.

ABSTRACT

Background: BKB SIN test is one of adaptive speech in noise tests. Its goal is to provide a mean of quantifying the real-word signal to noise ratio (SNR) loss that is not well inferred from the audiogram.
Aim: Develop and standardize the Arabic version of BKB SIN test sentences lists that measures the signal-to-noise ratio.
Patients and Methods: This study included 54 normal hearing children. Their age ranged from 5-18 years. Material: Sentences have words that are at first grade reading level and familiar to children. These sentences were recorded by male talker and were presented in four-talker babble in four phases.
Results: Phase I assess the equivalence of sentences and normalize signal to noise ratio 50%. In Phase II selection of sentences then construction of equivalent lists producing 18 equivalent lists, each list consists of eight sentences. The first sentence has four keywords, while the remaining sentences have three. Phase III determines the SNR starting level of lists. Phase IV provides age related norms. The final lists comprise the Arabic-BKB SIN test that measures the SNR a listener requires to understand 50% of keywords in sentences in a background of babble.
Conclusion: The Arabic-BKB test is developed and standardized. The statistical reliability and efficiency of the test suit it to practical applications.

Key Words: Bamford- Kowal- Bench speech in noise test (BKB SIN), signal to noise ratio loss (SNR loss) and children.

Received: 23 September 2020**, Accepted:** 12 November 2020

Corresponding Author: Amani Mohamed El-Gharib, MD, Department of Audio-vestibular Medicine, Faculty of Medicine, Tanta University, Egypt, **Tel.:** 01098469400, **E-mail**: amanielgharib@yahoo.com

ISSN: 2090-0740, March 2021 Vol.22, No.1

INTRODUCTION

Speech in noise (SIN) tests can directly address the most common complaint that patients have, which is an inability to hear well in background noise. Because it is a common complaint in all age ranges, (SIN) tests can provide valuable insights into what might be the most appropriate amplification strategy. The results of these tests give more precision in the way to counsel patients about realistic expectations.[1]

Children require a more advantageous signal to noise ratio (SNR) than adults to recognize speech in the presence of competing background sounds.[2] Children with hearing loss have greater difficulty recognizing speech in these situations when compared with age-matched peers with normal hearing.[3]

The BKB SIN was developed to overcome the shortcoming of the Quick SIN Test especially in young children. The IEEE (Institute of Electrical and Electronics Engineers) sentences in the Quick SIN are approximately at high school language level. These sentences are also lengthy which causes difficulties in testing cochlear implant (CI) users and adults with auditory memory deficits.[4]

There is no available standardized Arabic SIN tests suitable for children and cochlear implant (CI) users till now. Subsequently, the present work was aiming at development and standardization of Arabic version of BKB test to measure SNR loss.

PATIENTS AND METHODS:

This work was performed in the period between November 2017 and April 2019 at Audiovestibular Unit, ENT Department, Tanta University, with ethical approval code no. 31942/11/17. Communication was done with professor Killion to get the permission of developing the Arabic version of BKB test.

The study included 54 normal hearing children (25 males and 29 females). Their age ranged from 5-18 years. Normal hearing sensitivity and normal middle ear pressure.

 DOI: 10.21608/ejentas.2020.43910.1260

1. Equipment: Pure tone audiometry: Madsen Astera which is a Type-1, two channels, and PC-based audiometer with headphones of TDH39 type. Immittancemetry: Interacoustics (AT235) Impedance Audiometer. Zoom H1 Handy Recorder was used for recording of sentences and babble noise.

Arabic sentences that follow the English version of BKB test. Some of the sentences were chosen from Arabic HINT sentences.[5] While, some were chosen from children books at first grade reading level. The developed sentences have the following characteristics: (1) Short, highly redundant sentences. (2) Simple in semantic and syntactic context. (3) The vocabulary and grammar were familiar to children. (4) At approximately a first-grade reading level. (5) The sentences varied in length over a range that does not influence the ability of listeners to remember and repeat the entire sentence correctly. Sentence length was of three to five words. (6) Each sentence has three to four keywords. (7) The keywords were taken according to their frequency i.e. it must be a frequent word that often occurs in everyday speech. Recording of sentences was made by an Arabic native male speaker. He was instructed to maintain clarity, pace, and effort while reading the sentences. The sentences were recorded on Zoom H1 Handy Recorder and they were sampled as WAV to 24-bit/96KHz which used for high resolution recordings. Recordings were made in a double-walled sound treated room and a microphone was placed perpendicular to the speaker at a distance of 1 meter. Four Multitalker (three females and one male) babble was recorded by El-Rifaey *et al*.[6] Then, was transferred to computer programs and mixed with the recorded Arabic BKB sentences in a fashion that enables to direct separate inputs to the audiometer.

3. Procedure:

a. All subjects were submittedted to: Full audiological history, otological examination and basic audiological evaluation including: i.pure tone audiometry, ii.speech audiometry and iii.immittancemetry.

b. Development of Arabic Bamford- kowal- Bench speech in noise (BKB) test material through the following steps: Formation of sentence materials, recording of sentences by male talker, masking noise, formation of sentence babble combination and creation of lists.

c. Standardization of the normative data using Arabic BKB materials through multiple phases: At first a pilot study was done before starting the actual research in order to set the parameters that will be used. Sentences were presented monaurally at 70 dB HL to ten normal hearing children at -7, -2, +3 and +8dB SNR according to Etymotic Research.[7] The pilot study results revealed that -7dB was not very difficult and +8dB was very easy. Subsequently, sentences were tested for equivalence -7, -2 and +3dB SNR.

Phase I: 309 sentences were presented monaurally at 70 dB HL (loud but ok) to 15 children. Prior to the test session, 10 sentences were administered at 70 dB HL in a random manner to familiarize the children with the task. The 309 sentences were then presented three times (927 stimuli) at every SNR -7, -2 and +3dB SNR, in this order.[7] Due to the prolonged time of testing (one hour for each SNR), more than one session was needed and children were allowed to rest in between. Instructions: try to guess and repeat anything you think you heard the man say. Scoring: One point was given for each of five key words repeated correctly in each sentence. The SNR-50 is then calculated for each sentence using a formula based on Spearman-Kärber Equation for estimating threshold at the 50 % point of the psychometric function.[8]

Phase II: The purposes of phase II were (i) sentences selection was done using two selection criteria[7]: The first was that the standard deviation of the sentence should be 2 dB or less from the grand average value across sentences and subjects. The second was that the SNR-50 value for the sentence to be greater than -9.5 dB (values at or below –9.5dB indicated a test floor effect since they were below the reliable range for the -7, -2 and +3 dB SNRs used). (ii) Constructing of equivalent lists was done in phase II according to criteria provided by Etymoytic research.[7] The equivalence criterion was that the average value for each list does not deviate from the all-list grand average value by more than 1dB.

Phase III: 9 children with normal hearing participated in phase III. Stimuli: Equivalent lists were presented monaurally at 70 dB HL. These lists were arranged based on the SNR-50 values from phase II after application of equivalence criterion. Stimulus presentation: According to results of phase I, SNR 50% was (-5.48). So, in order to determine the starting level the equivalent lists were presented monaurally at 70 dB HL at SNR values of -2, +1, +4, +7, +10, +13, +16, +19 and +22dB in a special manner as shown in (Table 1).

Table 1: The order of lists presentation.

Subject	SNR (dB SNR)								
	-2	1	4	7	10	13	16	19	22
1	1,2	3,4	5,6	7,8	9,10	11,12	13,14	15,16	17,18
2	3,4	5,6	7,8	9,10	11,12	13,14	15,16	17,18	1,2
3	5,6	7,8	9,10	11,12	13,14	15,16	17,18	1,2	3,4
4	7,8	9,10	11,12	13,14	15,16	17,18	1,2	3,4	5,6
5	9,10	11,12	13,14	15,16	17,18	1,2	3,4	5,6	7,8
6	11,12	13,14	15,16	17,18	1,2	3,4	5,6	7,8	9,10
7	13,14	15,16	17,18	1,2	3,4	5,6	7,8	9,10	11,12
8	15,16	17,18	1,2	3,4	5,6	7,8	9,10	11,12	13,14
9	17,18	1,2	3,4	5,6	7,8	9,10	11,12	13,14	15,16

Phase IV: Stimuli: Formulated lists of eight sentences each, the first one has four keywords and the remaining sentences each have three, with one sentence at each of the following signal-to-noise ratios (3dB step): +13, +10, +7, +4, +1, -2, -5 and -8 dB. The level of the target talker remains constant and the level of the background babble increases by 3 dB for each sentence, to the 0 dB SNR level. In the last three sentences in these lists (sentences 6, 7 and 8) the level of the background babble remains constant (the same level as at the 0 dB SNR) and the level of the target talker decreases. This was done to reduce the SNR without increasing the relative constant overall presentation level. According to results of phase III +13 SNR was the starting level. Stimulus presentation: The formulated lists were presented monaurally at 70 dB HL to 30 normal hearing children. Scoring: One point was given for each key word repeated correctly in each sentence. The SNR-50 is then calculated for each list using a formula based on Spearman-Kärber Equation. The SNR steps were 3 dB starting at 13 dB SNR, so the starting point (13 dB) plus half of the step size (1.5), plus the extra word in the first sentence (1) equals to 15.5. SNR-50 equals to 15.5 minus the total number of words repeated correctly.

RESULTS:

Phase I included 15 normal hearing subjects with the mean age 16.4±1.1. Phase III included 9 normal hearing subjects with mean age 11.8±1.9 and phase IV included 30 normal hearing subjects divided into three age groups of ten participants each. The first group age ranged from 5 to 7 years with a mean±S.D of 5.67±0.54 years. The second group age ranged from above 7 to 10 years with a mean±S.D of 8.95±1.01 years. The third group age ranged from above 10 to 14 years with a mean±S.D of 12.6±1.17 years.

III.a. Results of phase I:

1. Characteristics of participants' response.

The mean of SNR 50 of 15 participants for 309 sentences was -5.53±2.75 ranging from -9.50 to 4.25 with standard error 0.21.

2. Normalization of the SNR values of the sentence-babble pairs: One point was given for each of five key words repeated correctly in each sentence. The SNR-50 for each sentence was calculated for using a formula based on the Spearman-Kärber Equation:[8]

SNR 50 = highest presentation level in dB SNR + ½ (step size in dB) – (step size in dB) (number of correct words) / (w)

w= the number of items per step size.

III.b. Results of phase II:

1- Sentences selection:

From total sentence number of 309 sentences, the remaining sentences were 151 sentences. The included sentence mean was -5.48 dB ranging from -9.5 to 1.04 dB.

2- Construction of equivalent lists:

Equivalence criterion was that the average value for each list could not deviate from the all-list grand average value by more than one dB. The final lists had a deviation from the grand list mean ranging from 0.02 up to 0.94. Included sentences were formed in 18 equivalent lists with 8 sentences of each list. One excluded sentence was added to the remaining 7 sentences to form a practice list.

III.c. Results of phase III:

1- Test the equivalence and homogeneity:

The mean of list pairs for each subject at 9 different SNR was calculated showing no significant difference denoting that the lists are equivalent. (*P value= 0.704*). Also, the mean of list pairs for subjects at each SNR was calculated showing significant difference denoting that the lists were relevant (*P value= 0.001*). Levene test was used to test the homogeneity of data of phase III and *p value = 0.039* indicated statistically significance for data homogeneity denoting good equalized lists.

2- Determination of starting SNR of lists:

The average of scoring at each SNR level is calculated to determine the starting SNR of the lists. The first SNR, at which the mean recognition rate = 100 % was chosen to be the starting level. Accordingly, the SNR + 13 was the starting level (Figure 1).

III. d. Results of phase IV:

- Normative data was divided into three subgroups according to language and vocabulary level.[7]

- List selection in each age group was done in two steps:

1. According to 95% confidence interval; any list with 95% confidence interval outside the 95% confidence interval of the mean of all lists was excluded.

2. Dispersion; the lists with dispersion above 75% percentile were excluded.

The first age group list selection and normative data

First step of list selection: showed in (Figure 2).

The second step of list selection:

The lists with dispersion above 75% percentile (lists 7, 11 and 18) were excluded (Figure 3).

The mean of SNR 50 of 10 participants for the final selected 12 lists was 4.12±1.36 ranging from -0.50 to 9.50.

The second age group list selection and normative data

First step of list selection: showed in (Figure 4).

Second Step of list selection:

The lists with dispersion above 75% percentile (lists 9, 13, 14 and 16) were excluded (Figure 5).

The mean of SNR 50 of 10 participants for the final selected 13 lists was -1.46±1.05 ranging from -6.50 to 2.50.

The third age group list selection and normative data

First step of list selection: showed in (Figure 6).

Second Step of list selection:

The lists with dispersion above 75% percentile (lists 8, 9 and 11) were excluded (Figure 7).

The mean of SNR 50 of 10 participants for the final selected 9 lists was -3.34±0.75 ranging from -7.50 to 0.50.

Results showed age related performance (Figure 8).

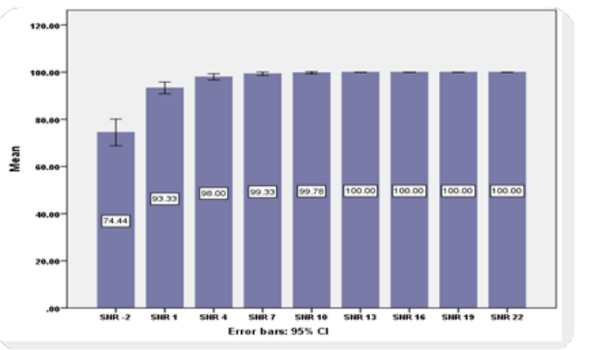

Fig. 1: The mean of subjects at each SNR level.

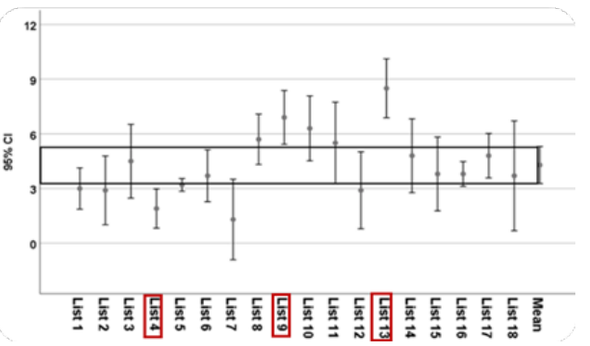

Fig. 2: Lists with their 95% CI of mean in the first age group.

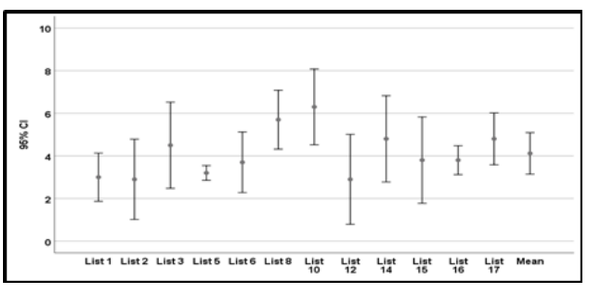

Fig. 3: The final lists with their 95% CI of mean in the first age group.

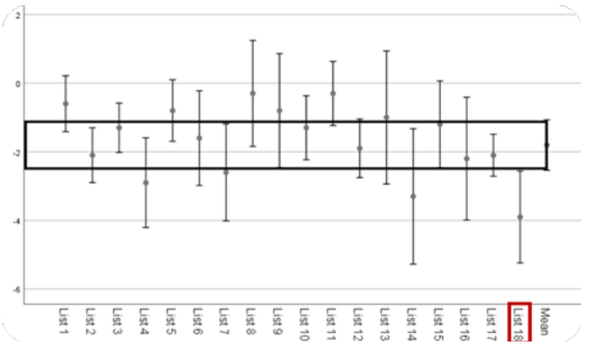

Fig. 4): Lists with their 95% CI of mean in the second age group.

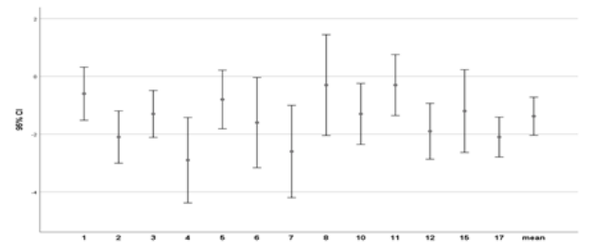

Fig. 5: The final lists with their 95% CI of mean in the second age group.

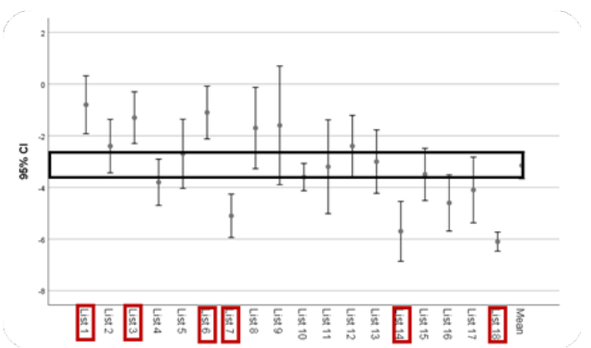

Fig. 6: Lists with their 95% CI of mean in the third age group.

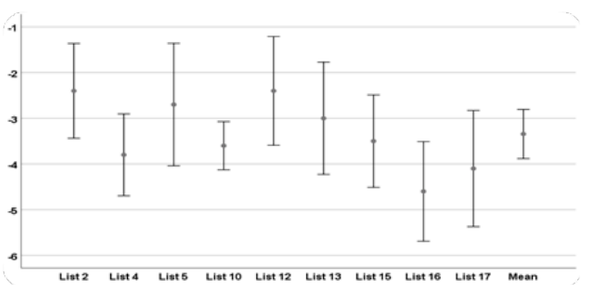

Fig. 7: The final lists with their 95% CI of mean in the third age group.

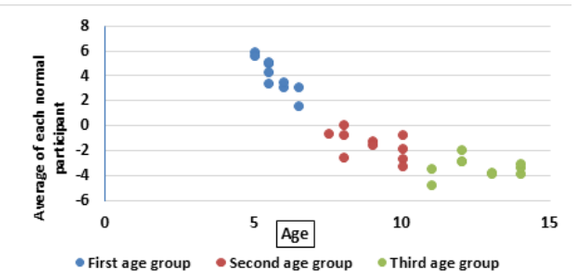

Fig. 8: The mean of SNR 50 in normal hearing children in different age groups.

Comparing one list, two list and three lists among 30 normal hearing subjects:

Analysis of variance (ANOVA) was done to compare the mean of the single list, paired lists, and the triple lists across the normal hearing subjects in the three age groups. No statistically significant difference was present (P value= 0.184, 0.286 and 0.858 in the first, second and third age group). Thus, The ANOVA test showed that using of single list is sufficient to obtain accurate SNR loss score.

Declarations

- The authors report no conflicts of interest with anybody.
- The authors declare they have no founding.
- The authors alone are responsible for the content and writing of the paper.

DISCUSSION

Results of phase I revealed SNR 50 values of each sentence-babble combination for each subject. According to results of this phase, sentences selection was done using two selection criteria as in Etymoytic research.[7] The first was that the standard deviation of the sentence should be 2 dB or less from the grand average value across sentences and subjects. The second was that the SNR-50 value for the sentence to be greater than -9.5 dB (values at or below –9.5dB indicated a test floor effect since they were below the reliable range for the -7, -2 and +3 dB SNRs used). These two criteria resulted in yielding 151 sentences and eliminating 158 sentences. While in English version 160 sentences were yielded at this phase. On the other hand, in the Persian version BKB test[9] the selection criteria depended on experts not on homogeneity and equivalence of sentences across normal subjects. This means that sentences were selected according to: content validity of the sentences through presenting them to 10 experts. And also standard deviation of SNR 50. Finally, sentences within the acceptable limit were chosen to be included and others were excluded.

SNR 50 of the included sentences was -5.48 dB which can be explained by the nature of test material (familiarity of words and grammar, short, highly redundant sentences rich with semantic and syntactic context and at approximately a first-grade reading level).

Constructing of equivalent lists was done in phase II according to criteria provided by Etymotic research.[7] The equivalence criterion was that the average value for each list does not deviate from the all-list grand average value by more than 1dB. Accordingly, this resulted in 18 equivalent lists, of eight sentences each. The remaining sentences were used as a practice list.

According to the results of pilot study, the starting level had to be readjusted. So, the main target of phase III was to search for the starting level. This was done by lists administration with several SNRs including: -2, +1, +4, +7, +10, +13, +16, +19 and +22dB (3 dB step) in a special manner as shown in the methodology (Table 1). This was done in order to reach the first SNR, at which the mean recognition rate equals to 100 % (starting level). This was done according to Zhou *et al*[10] who developed the Mandarin version of QuickSIN. The authors used this methodology in that test due to language difference between English and Chinese. In Arabic- BKB, the same methodology was used due to the same reason (difference between English and Chinese languages). In the current study, the starting level was +13 dB while in English BKB, it was +21 dB. That is because language difference between Arabic and English languages.

In order to test strength of the statistical power of lists, Levene test was used to test the homogeneity of data of phase III. Results revealed a p value of 0.039 indicating statistically significant homogeneity of data. This denoting that good equalized lists were used in phase III.

Moreover, ANOVA test was done, for comparing the mean of list pairs for each subject at the nine different SNRs. Results showed no significant difference which denoting that the lists are equivalent. Then ANOVA test was done again to compare the means of 9 list pairs at each SNR level. Results showed significant difference which denoting that the lists are relevant. These tests were not done in English nor Persian versions which make the Arabic BKB version more solid.

As the ability to understand speech in noise improves with age. In other words, the younger the age, the more SNR is needed. So, normative data had to be obtained across different ages. In addition, the information gained from normal hearing subjects will give an idea about the developmental trajectories of children. Also, it will verify that the test can discriminate between these developmental changes. Accordingly, this was done in phase IV. Normative data were divided into three age groups. These age groups were (5 to 7 years, >7 to 10 years and >10 to 14 years) according to exposure to language and language development. Results showed age related performance (Figure 8).

In phase IV, selection of lists for each age group was done using two selection steps: the first was the 95% confidence interval. Any list with 95% confidence interval outside the 95% confidence interval of the mean of all lists was excluded. The second selection step was the dispersion. Lists with dispersion above 75% percentile were excluded. This phase resulted in sensitive lists for each age group and age dependent normative data.

Eliminating lists in two selection steps for each age group help to reach the most sensitive lists for each age group. These selection steps ensured that the retained lists have low variability with high accuracy and reliability in testing children in these age groups.

Furthermore, the statistical power of Arabic BKB SIN lists was measured. In another words, whether a single list will be sufficient for accurate SNR loss measurement or not. So, analysis of variance (ANOVA) was done to compare the mean of the single lists, paired lists, and the triple lists across the normal hearing subjects. Results showed no statistically significant difference. This means that using a single list will be sufficient to obtain accurate SNR loss score. However, it is recommended to use triple lists especially in the first age group due to marked individual variability.

The English BKB test[7] provided age related norms while the Persian version[9] did not provide normal data for age groups. On the other hand, the current study developed age related norms as well as age related lists for each age group that will increase the ease of use as well as the sensitivity in each age group.

The Arabic-BKB SIN test used sentences spoken by a male talker in four-talker babble. It contains 18 single lists, eight sentences of each, with one sentence at each SNR (3 dB step) of: +13, +10, +7, +4, +1, -2, -5 and -8 dB. The first sentence in each list has four key words, and the remaining sentences have three keywords with sensitive lists in each age group. The key words in each sentence are scored as correct or incorrect. SNR 50 was calculated using Spearman Karber equation. Then results are compared to normative data to obtain the SNR loss.

CONCLUSION

Arabic BKB test can be used as a reliable and valid test for measuring SNR loss in children. The test is quick and easy to administer and score (about one to two minutes for each ear). The Arabic-BKB provides age-related norms for children. The Arabic-BKB provides sensitive lists for each age group (5-7, >7-10 and >10-14) (Table 2). Single list is sufficient to obtain accurate SNR loss score. However, we recommend the use of triple lists if time is available.

Moreover, we recommend application of Arabic BKB test on a large testing sample group of sensorineural hearing loss subjects. In addition, we recommend Integration of Arabic BKB test in the softwares of clinical audiometers especially that are sold in Middle East region.

Table 2: Lists no. for each age group.

Age group	No. of lists	Lists no. of this age group
5-7 years	12 lists	1, 2, 3, 5, 6, 8, 10, 12, 14, 15, 16, 17
>7-10 years	13 lists	1, 2, 3, 4, 5, 6, 7, 8, 10, 11, 12, 15, 17
>10-14 years	9 lists	2, 4, 5, 10, 12, 13, 15, 16, 17

CONFLICT OF INTEREST

There are no conflicts of interest.

REFERENCES

1. Taylor B. Using speech-in-noise tests to make better hearing aid selection decisions. 2011. Available at: http: //www .audiology online.com /articles /using-speech-in-noise-tests-832 [Accessed 20 May 2017].

2. Hall III J, Grose J, Buss E, Dev M. Spondee recognition in a two-talker masker and a speech-shaped noise masker in adults and children. Ear and Hearing. 2002; 23(2), pp.159-165.

3. Rance G, Barker E, Mok M. Speech perception in noise for children with auditory neuropathy/dys-synchrony type hearing loss. Ear Hear. 2007; 28, 351–360.

4. Sharma S, Tripathy R, Saxena U. Critical appraisal of speech in noise tests: a systematic review and survey. International Journal of Research in Medical Sciences. 2017; 5(1), pp.13-21.

5. Essawy W, Kabbash I, Kolkaila E, Emara A. Development and Standardization of Hearing in Noise test (HINT) in Arabic language. International Journal of otorhinolaryngology and head and neck surgery. 2019; vol.5, issue 6.

6. El rifaey M, El gharib A, El mahallawi T, Emara A . Development and Standardization of Arabic Quick Speech in Noise Test. Egyptian journal of ear nose throat and allied sciences. 2020.

7. Etymotic Research. Bamford-Kowal-Bench Speech-in-Noise Test (Version 1.03)-User manual. 2005. Retrieved from http://www.etymotic.com/ pdf/ bkbsin-user-manual.pdf.

8. Finney D. Statistical Method in Biological Essay. London. 1952. Quoted from (Wilson R H, McArdle R, Smith L . An Evaluation of the BKB-SIN, HINT, QuickSIN, and WIN Materials on Listeners with Normal Hearing and Listeners with Hearing Loss. Journal of Speech, Language, and Hearing Research. 2007; Vol. 50 • 844-856.

9. Moossavi A, Mehrkian S, Karami F, Biglarian A, Bakhtiari, B. Developing of Persian version of the BKB sentences and content validity assessment. Auditory and Vestibular Research. 2017; 26(1), pp.27-33.

10. Zhou R, Zhang H, Wang S, Chen J, Ren D. Development and Evaluation of the Mandarin Quick Speech-in-Noise Test Materials in Mainland China. J phonet Audiol. 2017; 2(1000124), p.2.