International Journal of Research in Medical Sciences
Sharma S et al. Int J Res Med Sci. 2017 Jan;5(1):13-21
www.msjonline.org
pISSN 2320-6071 | eISSN 2320-6012

DOI: http://dx.doi.org/10.18203/2320-6012.ijrms20164525

Review Article

Critical appraisal of speech in noise tests: a systematic review and survey

Suhani Sharma[1], Rajesh Tripathy[2], Udit Saxena[3]*

[1]Department of Audiology, National Speech and Hearing Centre, Hyderabad, India
[2]PhD student, Department of Psychology, Guwahati University, Assam, India
[3]MAA Institute of Speech and Hearing, MAA ENT Hospitals, 2nd Floor Block-A Shanti Sikara Complex, Raj Bhavan Road, Hyderabad, India, 500082

Received: 25 October 2016
Accepted: 26 November 2016

***Correspondence:**
Dr. Udit Saxena,
E-mail: dr.udit@maaent.com

ABSTRACT

Speech in noise tests that measure the perception of speech in presence of noise are now an important part of audiologic tests battery and hearing research as well. There are various tests available to estimate the perception of speech in presence of noise, for example, connected sentence test, hearing in noise test, words in noise, quick speech-in-noise test, bamford-kowal-bench speech-in-noise test, and listening in spatialized noise-sentences. All these tests are different in terms of target age, measure, procedure, speech material, noise, normative, etc. Because of the variety of tests available to estimate speech-in-noise abilities, audiologists often select tests based on their availability, ease to administer the test, time required in running the test, age of the patient, hearing status, type of hearing disorder and type of amplification device if using. A critical appraisal of these speech-in-noise tests is required for the evidence based selection and to be used in audiology clinics. In this article speech-in-noise tests were critically appraised for their conceptual model, measurement model, normatives, reliability, validity, responsiveness, item/instrument bias, respondent burden and administrative burden. Selection of a standard speech-in-noise test based on this critical appraisal will also allow an easy comparison of speech-in-noise ability of any hearing impaired individual or group across audiology clinics and research centers. This article also describes the survey which was done to grade the speech in noise tests on the various appraisal characteristics.

Keywords: Appraisal, Hearing, Noise, Speech perception, Survey

INTRODUCTION

Speech perception abilities get adversely affected in the presence of noise. Carhart et al highlighted the importance of estimating speech perception in noise in the regular auditory diagnostic battery.[1,2] Several speech-in-noise tests, for example, Connected Sentence Test (CST), Hearing in Noise test (HINT), Words in Noise (WIN), Quick Speech-in-noise Test (QuickSIN), Bamford-Kowal-Bench Speech in Noise Test (BKB-SIN), and Listening in Spatialized Noise-Sentences (LiSN-S), have been developed thereafter.[3-7] Audiologists may adapt to a speech-in-noise test depending on factors such as availability, ease to administer and time required to run the test, age of the patient, hearing status, type of hearing disorder, if the individual uses a hearing aid, etc. Although these factors are important considerations, a critical appraisal of speech-in-noise tests is necessary for making an evidence based selection.

This paper is aimed to critically appraise speech-in-noise tests (HINT, QuickSIN, BKB-SIN, LISN-S and WIN).

Characteristics described by Andersen to measure disability research outcomes were used to prepare this critical appraisal.[8] Those are Conceptual model, Measurement model, Norms, Reliability, Validity, Responsiveness, Item/Instrument bias, Respondent burden, Administrative burden, Alternate/accessible forms and Cultural/language adaptations.

CONCEPTUAL MODEL

Evaluating the conceptual model addresses the rationale for, and description of, the concepts that the test is intended to measure. It also addresses the relationships between those concepts.[9] Speech-in-Noise tests were developed to measure the ability to understand speech in presence of noise. Factors that may affect the speech-in-noise outcomes include speech materials, speaker of the speech material, type of noise and test set-up.[10] These factors are therefore necessary to consider while critically appraising the conceptual model of speech-in-noise tests. Speech stimuli used in speech-in-noise tests are phonemes, words or sentences. These stimuli have different language demands which can influence the test results. Intelligibility of speech in presence of noise is also dependent on the voice (male or female), type of noise (broadband, narrowband, speech-noise, speech-babble, etc) and the way test is administered (i.e. under headphones or in the sound-field).[11-14]

Hearing in noise test (HINT)

The HINT was developed by Nilsson et al for the measurement of Reception Threshold for Sentences (RTS) in quiet and in the presence of noise.[3] The goal of the HINT is to provide a reliable and efficient tool to estimate hearing handicap, directional hearing, hearing aid benefits and to perform comparison between hearing aids.[3,15]

The HINT includes 25 phonemically balanced lists of 10 sentences which were adapted from the Bamford-Kowal-Bench (BKB) sentences.[16] BKB sentences are suitable to test children above 4.6 years of age.[17] The BKB sentences included in the HINT were revised to remove British idioms and to equate the length of the sentences. In HINT, these sentences are recorded by male professional actor.[3] It is often documented that female voice is more intelligible than male but no justification has been provided by the developers of the HINT for using male voice.

The HINT uses a steady-state speech-shaped noise for measurement in presence of noise. This speech-shaped noise is spectrally matched to the long term average spectrum of the target sentence stimuli. The stability of level in speech-shaped noise can increase the reliability of individual SNR scores.[3] During the test noise is kept constant at 65 dB (A) while the intensity level of sentences changes in order to adapt to the dB SNR threshold. Original HINT was developed to measure speech in noise abilities in sound field. During the test sentences are always presented from 0° azimuth. For in-noise measurements, noise is presented either from 0°, 90° or 270° azimuth. Noise starts 100 msec before the sentence and ends 100 msec after the sentence is finished.[15] The HINT can also be used under headphones that can stimulate sound field.[3,15]

The HINT also has a children version; the HINT-C which was developed by Nillson et al.[18] The HINT-C has 13 lists of 10 sentences. Sentences used in HINT-C are from the subset of the HINT sentences that can be repeated correctly by 6 to 12 year old children. Use of male voice for sentences in the HINT is the only limitation in the conceptual model.

QuickSIN Test/ BKB-SIN Test

The QuickSIN Test, and the BKB-SIN Test, were developed to measure the signal-to-noise ratio loss in decibels (dB SNR loss).[5,6,9] SNR loss is described as the dB increase in signal-to-noise ratio required by an individual with hearing-impairment to understand speech in presence of noise as comparable to normal hearing individual.[19] The QuickSIN is a quick and improved version of the Speech-in-Noise (SIN) Test.[5] The QuickSIN test provide a quick measure of SNR loss, can quantify the benefits of directional microphones, and help the audiologist in choosing appropriate amplification options for individuals with hearing loss.[19] The QuickSIN is comprised of 12 lists, recorded by a female talker, each containing 6 sentences with 6 additional lists (3 lists for practice and another 3 for research). Each sentence consists of 5 keywords and each keyword is worth 1 point. Sentences used in the QuickSIN are adapted from Institute of Electrical and Electronics Engineers (IEEE) sentences which were later equalized to account for the high frequency attenuation present in the original recording from Massachusetts Institute of Technology.[20] These sentences provide limited contextual cues to the listeners. The QuickSIN uses 4-talker babble in place of noise in order to simulate real world listening condition.[5]

The BKB-SIN was developed to overcome the shortcoming of the QuickSIN Test especially in young children. IEEE sentences in the QuickSIN are approximately at high school language level.[6] These sentences are also lengthy which causes difficulties in testing cochlear implant (CI) users and adults with auditory memory deficits.[6] The BKB-SIN uses the BKB sentences. The BKB-SIN consist of 18 lists in which list number 1-8 has 10 sentences and from 9-18 has 8 sentences. First sentence of each list has 4 keywords and rest of the sentences have 3 keywords. Like the QuickSIN, the BKB-SIN also uses 4-talker babbles.

In both the tests, the presentation level of sentences is kept constant at 70 dB HL (83 dB SPL) and noise intensity decreases to change SNR. In the QuickSIN, multi-talker babble decreases in 5 dB steps such that SNR varies from +25 to +0 dB SNR. In the BKB-SIN, multi-talker babble increase in 3 dB step so that SNR varies from +21 to -6 dB SNR. Although 70 dB HL is the recommended presentation level for both tests, in the BKB-SIN presentation level can be reduced to 50 dB HL for children and CI users depending on their comfortable listening level. Both the QuickSIN and BKB-SIN can be done using headphones, binaural insert earphones or in sound field.

Listening in Spatialized Noise – Sentence (LiSN-S) Test

The LiSN-S test is a newer version of the original Listening in Spatialized Noise – Continuous Discourse test (LiSN-CD).[7,21] It was developed to assess speech-in-noise perceptual abilities in children as young as 5 years by incorporating a simplified and more objective response protocol.[7] Target sentences (120 sentences, female talker) used in LiSN-S are written by Australian Speech Language Pathologist. These sentences are constructed to be suitable for children from 4.6 years.[7] Stories are used as auditory distracter in LiSN-S. These stories are recorded by three female talkers (including the one who recorded target sentences). Both target and distracters are synthesized using Head Related Transfer Function (HRTF) so that the target sentences are perceived to be coming from 0° azimuth whereas the distracter stories (masker) vary according to the spatial location (0°, +90° and -90° azimuth), vocal identity of the speaker of story (same as, or different from the speaker of the target sentences) or both. This test configuration results in four distracter conditions: 1) same voice at 0° (SV0°); 2) same voice at +90° (SV+90°); 3) different voices at 0° (DV0°); and 4) different voices at +90° (DV+90°).[7] Performance on the LiSN-S is evaluated in terms of low and high-cue SRT and also on three “advantage” measures (Table 1). These advantage measures are the benefit in dB gained by cues like different talker (pitch), spatial cues and both talker and spatial cues compared to the low-cue SRT condition where no cues are present.7 There are 30 target sentences for all four conditions. Distracter stories are presented at a constant level of 55 dB SPL while the level of target sentences changes. The LiSN-S is administered under headphones.

Table 1: The LISN-S SRT and advantages measures (adopted from Cameron et al).

	Same speaker	Different speaker
Same location	Low cue SRT Same voice 0°	Talkers advantage Different voice 0°
Different location	Spatial advantage Same voice 0°	Total advantage High cue SRT Different voices

Words in Noise (WIN) Test

The WIN was developed by Wilson to measure the ability to understand speech in multitalker babble.[4] WIN is different from HINT, QuickSIN and LISN-S as it uses monosyllabic words instead of sentences like others. Wilson advocated for monosyllabic words as target speech material over sentences for speech-in-noise tests.[4] Sentences are not used widely in clinical audiology practice. Repeating sentence materials, especially in the presence of noise, involves many issues other than speech recognition for example recognition versus recall, memory, recency and primacy effects, bottom-up versus top-down information processing, and the multiplicative effects of various degradations on the recognition of speech signal. Also, the use of monosyllabic words makes the speech-in-noise measurements free of linguistic or contextual cues.

Monosyllabic words in WIN are adapted from the Northwestern University Auditory Test No. 6.4 Clinically available WIN materials contain two lists of 35 words (female talker). Like the QuickSIN, the WIN test also uses multi-talker babble (6 talker babble). The WIN test is typically administered under headphones. For each list in the WIN test, the noise level remains constant at 80 dB SPL. Intensity level of the speech material decreases in 4 dB step to vary SNR from +24 to 0 dB SNR. The WIN has a total of 7 SNR conditions and 5 monosyllabic words are presented at each SNR condition.

Table 2: Grading of HINT, QuickSIN, BKB-SIN, LiSN-S and WIN on conceptual model.

	Grade A	Grade B	Grade C
HINT	77.1	18.6	4.3
QuickSIN	86.2	12.8	1.1
BKB-SIN	66.0	28.2	5.9
LiSN-S	67.0	20.7	12.2
WIN	78.7	18.1	3.2

HINT: Hearing in noise test; QuickSIN: Quick speech in noise test; BKB-SIN: Bamford-Kowal-Bench speech in noise test; LiSN-S: Listening in spatialized noise-sentences; WIN: Word in noise

MEASUREMENT MODEL

The measurement model is reflected in the instrument’s scale and subscale structure and the procedures that are used to create the scale scores.[9] In simple words, the measurement model of an instrument should describe the procedure for obtaining the raw scores, methods adopted to convert these raw scores into a meaningful score (or value) that are comparable to the instrument’s scale, how that scale is developed and which psychometric procedure is followed. The HINT and LiSN use an adaptive method whereas the QuickSIN, BKB-SIN and WIN use method of constants to converge to their respective psychometric points. Both the methods have certain advantages and disadvantages. Adaptive procedures are more efficient, flexible and have less reliance on restrictive assumptions.[22] The major advantages of adaptive procedures are that they are highly efficient and as such they provide greater precision for a fixed number of observations. On the other hand, method of constants has the advantage of data being collected over a wide range. However if the test is aimed to estimate one point on the response curve then this method results in large number of observations that are placed away from the point of interest, thus are inefficient. Adaptive procedures have greater flexibility in terms of tracking the gradual drift in the measurement value or to compensate for the unpredictable trend in the measurement which is not possible with the method of constants.

Table 3: Grading of HINT, QuickSIN, BKB-SIN, LiSN-S and WIN on measurement model.

	Grade A	Grade B	Grade C
HINT	28.7	51.1	20.2
QuickSIN	33.5	44.7	21.8
BKB-SIN	53.2	33.5	13.3
LiSN-S	27.7	49.5	22.9
WIN	36.7	54.8	8.5

HINT: Hearing in noise test; QuickSIN: Quick speech in noise test; BKB-SIN: Bamford-Kowal-Bench speech in noise test; LiSN-S: Listening in spatialized noise-sentences; WIN: Word in noise

In addition the response curve derived from adaptive procedures is not always based on some parametric assumptions, e.g., presentation level of the stimulus, step size of the presentation level, etc. Experimenter can make changes to those parameters based on his knowledge and listener's age, listening state and responses. However adaptive procedures are complex and have a higher likelihood of eliciting a positive response as the level of stimulus increases. Sometimes adaptive procedure, especially in children, causes stress as signal approach threshold.

HINT & LISN-S

The HINT uses an adaptive procedure to converge at the 50% performance point for both in quiet and in noise measurements. The HINT follows a 100% criterion to be considered as the correct response for a sentence with some flexibility given in responses for articles. For both the conditions (quiet and noise), an initial ascending approach is used to determine the presentation level at which the first sentence is correctly repeated.

The presentation levels of sentences 2 through sentence 4 are adaptively increased (for an incorrect response) or decreased (for a correct response) in 4 dB steps. The presentation levels of sentences beginning with sentence 5 are adaptively increased or decreased in 2-dB steps. RTS is then calculated by averaging the presentation level of the last 7 sentences. In noise condition, noise is presented at a constant level of 65 dB (A). This noise level, when deducted from RTS measured in the noise condition, gives the dB SNR required to reach 50% correct performance.[3,15] Percent intelligibility, measured at a fixed level of signal and/or noise, may be prone to floor and ceiling effects.[3] For this reason the developers of the HINT adopted an adaptive method for the HINT to eliminate floor and ceiling effects.

Like the HINT, the LiSN-S is an adaptive test. For each condition SNR is adapted by increasing or decreasing the target level in 4 dB steps until the first reversal and then in 2 dB step to determine individual SRT in noise (dB SNR). SNR decrease by 2 dB if a listener repeats more than 50% of the words correctly, increase by 2 dB if listener score less than 50% and SNR remain unchanged if the response is exactly 50% correct. Testing stops it if listener completes the entire 30 sentences in the list or completes the practice sentences with a minimum of 17 target sentences with an standard error (which is calculated automatically in real time over the scored sentences) is less than 1 dB.[23] This is called standard error (SE) stopping rule. The formula for SE stopping rule for LiSN-S is a modification of SE stopping rule for Monte Carlo simulation of an adaptive test sequence.[7]

Threshold is the mean of the stimulus level used in the measurement. A method to check the reliability of the listener's threshold is to calculate the SE of the stimulus levels used in determining the threshold. The goal of the test is to obtain thresholds with particular accuracy and for adaptive tests it is possible by stopping the test when SE decreases below a specified value.[7] SE stopping rule aids software based tests to test all the participants to the same level of precision irrespective of the number of items each required to reach the specified SE (Lilley, Barker and Britton.[24] Although HINT and LiSN-S are precise and flexible, they do rely on a presumed parameter which is unlike typical adaptive tests. This occurs because these tests comes with preset presentation level and step size and the clinician doesn't have control over the presentation level of the stimulus as well as step size of varying SNR.

QuickSIN, BKB-SIN and WIN

QuickSIN, BKB-SIN and WIN uses a descending level paradigm (method of constants) to calculate the correct score. This is followed by calculating 50% correct performance using Spearman Karber. The Speraman Karber formula is:[10]

$$\text{50\% correct performance} = \frac{i + \tfrac{1}{2}\,(d) - (d)\,(\text{correct score})}{w}$$

In which i = the initial SNR, d = step size of presentation level, and w = the number of items at each level.

Like HINT, QuickSIN, BKB-SIN and WIN tests are designed to be kept free from floor and ceiling effects by avoiding calculating percent intelligibility at fixed signal and/or noise level. The 50% correct point in QuickSIN (SNR 50) is calculated by subtracting total correct words from 27.5 i.e "27.5 - Correct score" (simplified Spearman

Karber equation for QuickSIN).[19] For BKB-SIN simplified equation is “27.5 - Correct score”. Similarly in WIN, the 50% correct point is calculated from the equation “26 - Correct score (out of 35 wordss) * (0.8)” (simplified Spearman Karber equation for WIN). In addition, the QuickSIN and BKB-SIN also measures SNR loss which is calculated by subtracting listener's SNR-50 from average SNR-50 of normal listeners.

Further, QuickSIN, BKB-SIN and WIN have their respective classification to categorize the speech-in-noise abilities. Classification of QuickSIN and BKN-SIN based on SNR loss: 0 – 3 dB SNR (near normal hearing), 3 – 7 dB SNR (mild SNR loss), 7 – 15 dB SNR (moderate SNR loss) and >15 dB SNR (severe SNR loss). The WIN classifies speech-in-noise abilities depending on SNR-50 threshold: normal (-2 to 6 dB SNR), mild loss (6.8 to 10 dB SNR), moderate loss (10.8 to 14.8 dB SNR), severe loss (15.6 to 19.6 dB SNR) and profound loss (20.4 to 25.2 dB SNR).[6,19]

Like any method of constants procedure followed in the QuickSIN, BKB-SIN and WIN has the same disadvantage in terms of inefficiency as it takes a large number of trials in order to collect the data and these points are poorly placed relative to the desired point on psychometric function (50% point in these tests). In addition, this method involves measures at very high and low SNR that may introduce floor and ceiling effects in the determination of the threshold (dB SNR) or SNR loss.

NORMS AND STANDARD VALUES

Normative data of clinical test allows comparison of an individual test result's to that of the normal population. The CD version of the HINT has normative data for adults.[15]

Table 4: Grading of HINT, QuickSIN, BKB-SIN, LiSN-S and WIN on normative.

	Grade A	Grade B	Grade C
HINT	82.4	13.8	3.7
QuickSIN	54.3	31.4	14.4
BKB-SIN	53.2	25.5	21.3
LiSN-S	39.9	32.4	27.7
WIN	63.3	18.6	18.1

HINT: Hearing in Noise Test; QuickSIN: Quick Speech in Noise Test; BKB-SIN: Bamford-Kowal-Bench Speech in Noise Test; LiSN-S: Listening in Spatialized Noise-Sentences; WIN: Word in Noise

Normative data for children are also available in HINT – C.[18] HINT has been used extensively in experimental and clinical research and as such a good amount of comparative data is present. As a measure of speech recognition in noise HINT is used in several experiments to evaluate directional microphone performance, in comparative studies, including comparison of different hearing aids and for determining the benefits from a particular hearing aid or cochlear implant.[25-27] HINT is also used extensively for speech in noise measure in clinical population like presbycusis, sloping hearing loss, central auditory processing disorder (CAPD), and auditory neuropathy.[29-33]

QuickSIN test has normative data only on adults.[5,19] Some comparative data on QuickSIN is available from studies including Bochner et al, Duncan et al, McArdle et al and Wilson et al.[20,34-36] Compared to the HINT, the QuickSIN test has been used very little in peer reviewed studies. Although the QuickSIN has been developed to measure the benefits from hearing aids and other devices, the literature shows very few such studies. BKB-SIN test has normative data on children from 5 to 14 years and also on adults.[6] Being recent developed test very sparse comparative date is available on BKB-SIN.[36,37]

Normative data on LiSN-S from 5 to 11 years is published in Cameroon et al.[7] Normative data on LiSN-S is also available on adolescents and adults up to 60 years of age.[38] Comparative data of children with CAPD and learning disability (LD) on LiSN-S is present in Cameron et al.[39] The original article by Wilson has normative scores for normal hearing young adults.[4] Normative data on children from 6 to 12 years are also available.40 Comparative scores in hearing impaired listeners on WIN are also published in several studies.[41-43]

RELIABILITY

Reliability of the test refers to the consistency of its measure.[44] Reliability can be estimated in two ways; 1) for internal consistency, that is, high correlations among test items, and 2) for reproducibility, that is, the stability over time in test-retest circumstances.[9] The original article on the development of HINT evaluated the reliability of RTS measured with different sentence lists of HINT.[3] They calculated the reliability of RTS in quiet and noise condition based on the standard deviation of differences in repeated RTS measures within subjects. The mean RTS in both quiet and noise conditions for all lists was within 1 dB of the overall mean. The standard deviation of the difference score was 1.39 dB in quiet and 1.13 dB in noise.

Table 5: Grading of HINT, QuickSIN, BKB-SIN, LiSN-S and WIN on reliability.

	Grade A	Grade B	Grade C
HINT	78.2	19.1	2.7
QuickSIN	80.9	15.4	3.7
BKB-SIN	68.6	20.2	11.2
LiSN-S	70.7	21.8	7.4
WIN	78.7	11.2	10.1

HINT: Hearing in Noise Test; QuickSIN: Quick Speech in Noise Test; BKB-SIN: Bamford-Kowal-Bench Speech in Noise Test; LiSN-S: Listening in Spatialized Noise-Sentences; WIN: Word in Noise

On the basis of these results authors concluded that reliable RTS in quiet and noise can be obtained from HINT. Killion et al described the reliability of QuickSIN on the basis of number of lists required to produce a given accuracy in hearing impaired subjects at the 80% and 95% confidence interval.[5]

Results indicated that the reliability of a QuickSIN score depends on the number of lists used in the test. For example, at the 95% confidence interval, a QuickSIN score for 3 lists is accurate to + 1.3 dB whereas the accuracy increased to + 0.7 dB for 9 lists. Similar to QuickSIN, reliability of the BKB-SIN is described in terms of number of lists required to reach accuracy at 95% and 80% confidence interval for adults, CI users and children at different age ranges.[6] Cameron et al, measured inter-session test-retest reliability on the LiSN-S in 130 adolescents and adults.[38] They revealed that there was a small but significant test–retest improvement (between 0.5 and 1.2 dB) across all age ranges on the four LiSN-S conditions, but there was no significant test-retest difference in scores on the spatial, talker, or total advantage measures.

Wilson et al investigated test- reliability within session and between sessions (separated by 2-3 months) in 48 listeners with mild to severe hearing loss and 48 with moderate to severe hearing loss.[45] Results in both groups indicated no significant differences in WIN scores for List 1 and List 2. There was a slight improvement in performance from Trial 1 to Trial 2, but it was not significant and this finding indicates good intra-session reliability (the intra-class correlation for the test-retest data were 0.89 for the mild hearing loss group and 0.91 for the moderate hearing loss group). Further, no significant intra-session performance differences were found. Wilson also reported good intra-session test retest reliability in 24 normal hearing listeners.[4]

VALIDITY

Validity is the measure of the degree of accuracy with which the test measure, what it is designed to measure. In other words, it is the reflection of accuracy in measurements for what the test was intended to measure.

Table 6: Grading of HINT, QuickSIN, BKB-SIN, LiSN-S and WIN on validity.

	Grade A	Grade B	Grade C
HINT	68.1	22.3	9.6
QuickSIN	70.7	16.5	12.8
BKB-SIN	61.2	25.0	13.8
LiSN-S	69.7	19.1	11.2
WIN	77.1	15.4	7.4

HINT: Hearing in Noise Test; QuickSIN: Quick Speech in Noise Test; BKB-SIN: Bamford-Kowal-Bench Speech in Noise Test; LiSN-S: Listening in Spatialized Noise-Sentences; WIN: Word in Noise

Validity of a test is therefore different from its reliability, although reliability is a prerequisite to validate a test. Similar results on any test across different studies on a particular population (e.g. same age, status of hearing, hearing disorder, etc.) reflect good validity of the tests. The various speeches in noise tests discussed in this paper have shown similar results across different studies when testing same populations. Therefore all the tests are expected to have validity.

RESPONSIVENESS

Tests with good responsiveness are able to accurately detect a change in performance whether it is clinically significant or not.[8] The HINT and LiSN-S are adaptive tests where SNR changes depending on patient response. An increase or decrease in noise level is reflected by the change in the response from the listener in terms of correct or incorrect response. Also, these tests include spatial change of the noise source which is detected in a similar manner, such as a similar location of signal and noise source result in poor performance compared to when they are spatially separated. Additionally, the LiSN-S also includes conditions like same and different speaker for target and distracter sentences and this is detected in the form of better responses in different speaker presentation (with pitch as a cue).

Table 7: Grading of HINT, QuickSIN, BKB-SIN, LiSN-S and WIN on responsiveness.

	Grade A	Grade B	Grade C
HINT	50.5	35.6	13.8
QuickSIN	78.7	12.8	8.5
BKB-SIN	68.6	17.6	13.8
LiSN-S	55.9	27.1	17.0
WIN	77.1	19.1	3.7

HINT: Hearing in Noise Test; QuickSIN: Quick Speech in Noise Test; BKB-SIN: Bamford-Kowal-Bench Speech in Noise Test; LiSN-S: Listening in Spatialized Noise-Sentences; WIN: Word in Noise

In QuickSIN, BKB-SIN and WIN, SNR decreases as the test progresses which are represented by low score every reduction in SNR. These are the only predictable changes in listener’s response depending on the construct of these tests and can be easily detected.

ITEM/INSTRUMENT BIAS

All the tests have pre-test criteria for the speech and noise levels. Also pre-determined is the manner and size of the changing SNR during the test. Normative data on all these tests are based on the steps suggested in the manual/published article on the test development. When provided these tests are performed following the suggested instructions, including all the lists and conditions given in the original version, they are not biased for any population provided the normative or standard comparative values of the test are available for that population and language.

Table 8: Grading of HINT, QuickSIN, BKB-SIN, LiSN-S and WIN on item/instrument bias.

	Grade A	Grade B	Grade C
HINT	54.8	36.2	9.0
QuickSIN	51.1	38.3	10.6
BKB-SIN	59.6	34.6	5.9
LiSN-S	52.1	43.6	4.3
WIN	67.0	28.7	4.3

HINT: Hearing in Noise Test; QuickSIN: Quick Speech in Noise Test; BKB-SIN: Bamford-Kowal-Bench Speech in Noise Test; LiSN-S: Listening in Spatialized Noise-Sentences; WIN: Word in Noise

RESPONDENT BURDEN

On the basis of listener response

HINT, QuickSIN, and WIN are designed to measure 50% correct performance. In HINT, correct response of a sentence is based on 100% correct repetition by the listener, which makes it difficult especially for the children. In contrast, the QuickSIN and WIN tests are scored depending on the number of target words (5 words) correctly repeated for the sentence. The LiSN-S also has a 50% criterion for correct response for a sentence. On the basis of these characteristics, respondent burden is greater in the HINT as compared to the QuickSIN, BKB-SIN, LISN-S and WIN tests.

On the basis of speech material

WIN uses monosyllabic words whereas HINT, QuickSIN, BKB-SIN and LiSN uses sentence as speech materials. Linguistic cues available in sentence material makes the speech recognition tasks comparatively easy when compared to monosyllabic words. This increases the respondent burden in WIN as compared to other tests. Further, the HINT, BKB-SIN and LiSN has sentence material suitable for age from 4.6 years, so these are relatively easier for children when compared to QuickSIN. The HINT-C in addition has equivalent sentences from the HINT that could be easily repeated by children from 5- to 6- years.

On the basis of the time taken

HINT consists of 25 lists of 10 sentences; QuickSIN has 12 lists of 6 sentences; BKB-SIN has sentence lists of 8 – 10 sentences; LiSN-S has 180 sentences to be tested in 4 conditions; and WIN has 2 lists of 35 wordss presented in 4 random orders. As such all the test are lengthy and require quite a bit of time to complete. Based on test time the respondent burden for an individual undergoing speech-in-noise testing is much more than for other behavioural tests.

On the basis of measurement model

As discussed in the section of measurement model, the respondent burden on adults is the same regardless of the use of an adaptive procedure (HINT and LiSN) or method of constants (QuickSIN, BKB-SIN and WIN). Children might be less stressed in the test following a method of constants.

Table 9: Grading of HINT, QuickSIN, BKB-SIN, LiSN-S and WIN on respondent burden.

	Grade A	Grade B	Grade C
HINT	18.1	59.6	22.3
QuickSIN	71.3	19.1	9.6
BKB-SIN	68.1	23.9	8.0
LiSN-S	29.3	54.3	16.5
WIN	58.0	36.2	5.9

HINT: Hearing in Noise Test; QuickSIN: Quick Speech in Noise Test; BKB-SIN: Bamford-Kowal-Bench Speech in Noise Test; LiSN-S: Listening in Spatialized Noise-Sentences; WIN: Word in Noise

ADMINISTRATIVE BURDEN

HINT has both compact disc (CD) version, and software version.[3] In CD version the administrator has to change the levels of the sound depending on the listener's response, write down the stimulus presentation level for each sentence, has to calculate the average of the presentation level of last 7 sentences in the list of 10 to calculate RTS and finally, administer has to calculate dB SNR in different conditions. In software version the administrator has to enter whether the response is correct or not. The software calculates the dB SNR by itself. LISN-S is available only in software. Similar to software based HINT the administrator only enters the response. Based on the purpose of the test the administrator compares dB SNR obtained in different conditions of the test. QuickSIN, BKB-SIN and WIN tests are available on CD. SNR for different lists changes by itself on the completion of the list. Administer have to calculate the correct response scores and put them to formula to derive dB SNR.

Table 10: Grading of HINT, QuickSIN, BKB-SIN, LiSN-S and WIN on administrative burden.

	Grade A	Grade B	Grade C
HINT	71.3	20.7	8.0
QuickSIN	77.7	16.5	5.9
BKB-SIN	70.2	19.7	10.1
LiSN-S	68.6	23.9	7.4
WIN	58.0	36.2	5.9

HINT: Hearing in Noise Test; QuickSIN: Quick Speech in Noise Test; BKB-SIN: Bamford-Kowal-Bench Speech in Noise Test; LiSN-S: Listening in Spatialized Noise-Sentences; WIN: Word in Noise

ALTERNATE/ACCESSIBLE FORMS

HINT, QuickSIN, BKB-SIN, LiSN-S and WIN have no other alternate form available for special groups.

CULTURAL/LANGUAGE ADAPTATIONS

HINT has been developed in several other language like Swedish, Latin-American, Canadian-French, Mandarin, etc. LISN-S has two versions i.e. Australian version and North American version. QuickSIN, BKB-SIN and WIN have only one version.

SURVEY TO GRADE SPEECH IN NOISE TESTS FOR APPRAISAL CHARACTERISTICS

A small survey was conducted to grade the HINT, QuickSIN, BKB-SIN, LiSN-S and WIN on the characteristics reviewed in this article; conceptual model, measurement model, norms, reliability, validity, responsiveness, item/instrument bias, respondent burden and administrative burden. As most of these tests are available in CD/software and only in English language, alternate/accessible forms and cultural/language adaptations were not included in the survey. 251 audiology professional consisting of 113 post graduate audiology students, 32 academicians, 18 researchers and 88 clinical audiologists were contacted through emails for this survey. They were asked to grade the speech in noise tests on the various characteristics using the grading criterion similar to what was used by Andersen.[8] Three grades (A, B and C) were used; A: Good, B: Average & C: Poor. Explanation of the various appraisal characteristics was sent electronically along with the grading sheet to the participants. One eighty eight professional revert back out of 251 to whom the survey was sent. Grades obtained from all the respondents were used to calculate percentage (of choice) for grades for each test and its appraisal characteristic. Results for appraisal characteristic; conceptual model, measurement model, norms, reliability, validity, responsiveness, item/instrument bias, respondent burden and administrative burden are shown in Table 1-9 respectively. Results are represented in terms of percentage which indicates the number of respondents (out of 188) who choose the respective grade (A, B or C) for a particular test and its appraisal characteristic.

CONCLUSION

The critical appraisal done in this article should help the audiologists in making an evidence based selection of speech in noise test in their practice. The various speech in noise tests are constructed with the aim of testing the same domain of hearing but are different in terms of the age they are specifically made for, type of speech stimuli used in the test, type of noise, etc. Such are the factors which make audiologist’s choice variable for selecting speech in noise test. This variability was also shown in the result of survey’s grading. It was seen that that for most of the critical appraisal characteristics Grade A was the most common choice followed by B while Grade C was given rarely.

Funding: No funding sources
Conflict of interest: None declared
Ethical approval: Not required

REFERENCES

1. Carhart R, Tillman TW. Interaction of competing speech signals with hearing losses. Acta Oto-Laryngol. 1970;91:274-9.
2. Cox RM, Alexander GC, Gilmore C. Development of the Connected Speech Test. Ear Hear. 1987;8:119S-25S.
3. Nilsson M, Soli SD, Sullivan JA. Development of the Hearing In Noise Test for the measurement of speech reception thresholds in quiet and in noise. J Acoust Soc Am. 1994;95(2):1085-99.
4. Wilson RH. Development of a Speech-in-Multitalker - Babble Paradigm to Assess Words-Recognition Performance. J Am Acad Audiol. 2003;14(9):453-70.
5. Killion MC, Niquette PA, Gudmundsen GI, Revit LJ, Banerjee S. Development of a quick speech-in-noise test for measuring signal-to-noise ratio loss in normal-hearing and hearing-impaired listeners. J Acoust Soc Am. 2004;116(4):2395-405.
6. Etymotic Research. Bamford-Kowal-Bench Speech-in-Noise Test (Version 1.03)-User manual; 2005. Retrieved fromhttp://www.etymotic.com/ pdf/ bkbsin-user-manual.pdf
7. Cameron S, Dillon H. Development of the Listening in Spatialized Noise-Sentences Test (LISN-S). Ear Hear. 2007;28(2):196-211.
8. Andresen EM. Criteria for Assessing the Tools of Disability Outcomes Research. Arch Phys Med Rehabil. 2000;81(2):15-20.
9. Lohr KN, Aaronson NK, Alonso J, Burnam MA, Patrick DL, Perrin EB, et al. Evaluating quality of life and health status instruments. Development of scientific review criteria. ClinTher. 1996;18:979-92.
10. Wilson RH, McArdle RA, Smith SL. An Evaluation of the BKB-SIN, HINT, QuickSIN, and WIN Materials on Listeners With Normal Hearing and Listeners With Hearing Loss. J Speech Lang Hear Res. 2007;50:844-57.
11. Ferguson SH. Talker differences in clear and conversational speech: Vowel intelligibility for normal-hearing listeners. J Acoust Soc Am. 2004;116(4):2365-73.
12. Bradlow AR, Bent T. The clear speech effect for non-native listeners. J Acoust Soc Am. 2002;112(2):271-83.
13. Carhart R, Tillman T, Greetis R. Perceptual masking in multiple sound backgrounds. J Acoust Soc Am. 1969;45,694-703.
14. Hall JW, Grose JH, Buss E, Dev MB. Spondee Recognition in a Two-Talker Masker and a Speech-Shaped Noise Masker in Adults and Children. Ear Hear. 2002;23(2):159-65.
15. House Ear Institute. Hearing in Noise Test operator's manual. Los Angeles, CA: Starkey Laboratories. 1995.

16. Bench J, Bamford J. Eds. Speech-hearing tests and the spoken language of hearing-impaired children. Academic Press. 1979.
17. Bench J, Kowal A, Bamford J. The BKB (Bamford-Kowal-Bench) sentence lists for partially-hearing children. Br J Audiol. 1979;13(3):108-12.
18. Nilsson MJ, Soli SD, Gelnett DJ. Development and norming of Hearing In Noise test for children. House Ear Institute (Internal report). 1996.
19. Etymotic Research. Quick Speech-in-Noise Test (Version 1.3)-User manual; 2001. Retrieved fromhttp://www.etymotic.com /pdf/ quicksin-manual.pdf
20. Duncan KR, Aarts NLA. Comparison of the HINT and QuickSIN Tests. J Speech Lang Pathol Audiol. 2006;30(2):86-94.
21. Cameron S, Dillon H, Newall P. Development and Evaluation of the Listening in Spatialized Noise Test. Ear Hear. 2006;27(1):30-42.
22. Levitt HCCH. Transformed up-down methods in psychoacoustics. J Acoust Soc Am. 1971;49(2B):467-77.
23. Cameron S, Dillon H. The Listening in Spatialized Noise–Sentences Test (LISN-S): Comparison to the Prototype LISN-S and Results from Children with Either a Suspected (Central) Auditory Processing Disorder or a Confirmed Language Disorder. J Am Acad Audiol. 2008;19:377-91.
24. Lilley M, Barker T, Britton C. The development and evaluation of a software prototype for computer-adaptive testing. Computers & Education. 2004;43(1):109-23.
25. Valente M, Schuchman G, Potts LG, Beck LB. Performance of dual microphone in-the-ear hearing aids. J Am Acad Audiol. 2000;11:181-9.
26. Cord ML, Leek MR, Walden BE. Speech recognition ability in noise and its relationship to perceived hearing aid benefit. J Am Acad Audiol. 2000;21:75-483.
27. Tyler RS, Gantz MI, Rubenstein IT, Wilson BS, Parkinson AJ, Wolaver A, et al. Three-month results with bilateral cochlear implants. Ear Hear 2002;23:80S-9S.
28. Humes LE, Hailing D, Coughlin M. Reliability and stability of various hearing-aid outcome measures in a group of elderly hearing-aid wearers. J Speech Lang Hear Res. 1996;39:923-35.
29. Saunders GH, Cienkowski KM. A Test to Measure Subjective and Objective Speech Intelligibility. J Am Acad Audiol. 2002;13(1):38-49.
30. Walden TC, Walden BE, Cord MT. Performance of custom-fit versus fixed-format hearing aids for precipitously sloping high-frequency hearing loss. J Am Acad Audiol. 2002;13:356-66.
31. Fuente A, Mcpherson B, Munoz V, Espina JP. Assessment of central auditory processing in a group of workers exposed to solvents. Acta Oto-Laryngol. 2006;126:1188-94.
32. Johnston KN, John AB, Kreisman NV, Hall JW, Crandell CC. Multiple benefits of personal FM system use by children with auditory processing disorder (APD). Int J Audiol. 2009;48:371-83.
33. Peterson A, Shallop J, Driscoll C, Breneman A, Babb J, Stoeckel R, et al. Outcomes of Cochlear Implantation in Children with Auditory Neuropathy. J Am Acad Audiol. 2003;14(4):188-201.
34. Bochner JH, Garrison WM, Sussman JE, Burkard RF. Development of Materials for the Clinical Assessment of Speech Recognition : The Speech Sound Pattern Discrimination Test. J Speech Lang Hear Res. 2003;46:889-901.
35. McArdle RA, Wilson RH Burks CA. Speech Recognition in Multitalker Babble Using Digits, Words, and Sentences. J Am Acad Audiol. 2005;16(9):726-39.
36. Wilson RH, McArdle RA Smith SL. An Evaluation of the BKB-SIN, HINT, QuickSIN, and WIN Materials on Listeners with Normal Hearing and Listeners with Hearing Loss. J Speech Lang Hear Res. 2007;50:844-57.
37. Ng SL, Meston CN, Scollie SD, Seewald RC. Adaptation of the BKB-SIN test for use as a pediatric aided outcome measure. J Am Acad Audiol. 2011;22:375-86.
38. Cameron S, Glyde H, Dillon. Listening in Spatialized Noise—Sentences Test (LiSN-S): Normative and Retest Reliability Data for Adolescents and Adults up to 60 Years of Age. J Am Acad Audiol. 2011;22(10):697-709.
39. Cameron S, Dillon H. The Listening in Spatialized Noise–Sentences Test (LISN-S): Comparison to the Prototype LISN-S and Results from Children with Either a Suspected (Central) Auditory Processing Disorder or a Confirmed Language Disorder. J Am Acad Audiol. 2008;19:377-91.
40. Wilson RH, Farmer NM, Gandhi A, Shelburne E, Weaver J. Normative data for the Words-in-Noise Test for 6- to 12-year-old children. J Speech Lang Hear Res. 2010;53(5):1111-21.
41. McArdle RA, Wilson RH, Burks CA. Speech Recognition in Multitalker Babble Using Digits, Words, and Sentences. J Am Acad Audiol. 2005;16(9):726-39.
42. Wilson RH, Cates WB. A Comparison of Two Words-Recognition Tasks in Multitalker Babble: Speech Recognition in Noise Test (SPRINT) and Words-in-Noise Test (WIN). J Am Acad Audiol. 2008;19(7):548-56.
43. Wilson RH, Weakley DG. The use of digit triplets to evaluate words-recognition abilities in multitalker babble. Seminars in Hearing. 2004;25:93-111.
44. Sechrest L. Reliability and Validity. Research Methods in Clinical Psychology; 1984. Retrieved from http://www.egadconnection.org/ Reliability%20and%20validity.pdf.
45. Wilson RH, McArdle R. Intra- and Inter-session Test, Retest Reliability of the Words-in-Noise (WIN) Test. J Am Acad Audiol. 2007;18(10):813-25.

Cite this article as: Sharma S, Tripathy R, Saxena U. Critical appraisal of speech in noise tests: a systematic review and survey. Int J Res Med Sci 2017;5:13-21.