Speech Perception in Noise Measures for Children: A Critical Review and Case Studies

Erin C. Schafer, Ph.D.
University of North Texas
Denton, Texas

Children who have hearing loss or other auditory disorders are at risk for educational difficulties, especially when the detrimental effects of an impaired auditory system are combined with poor classroom acoustics. Classroom observations, teacher questionnaires, and speech perception measures in noise may be used to identify children who are at-risk and to evaluate the effects of classroom noise on behavior and performance. Valid and reliable quantification of listening difficulty will provide evidence of a child's need for instructional and communication accommodations, special education support, and hearing assistance technology. Currently, however, no cumulative peer-reviewed publications that analyze speech perception tests in noise for children exist. For this reason, the primary goal of this paper is to provide a critical review of speech perception measures in noise which are designed for young and school-aged children. This review will provide information regarding the sensitivity, validity, and reliability of available measures, as well as discuss advantages and disadvantages of each test for examining pediatric speech perception in noise. In addition, three case studies will demonstrate the clinical utility of two tests for measuring speech perception in noise in a classroom setting.

Introduction

One of the greatest challenges for audiologists is identifying and addressing the deleterious effects of classroom noise and reverberation on speech perception of children who have hearing loss and other auditory disorders (Bradley & Sato, 2008; Jamieson, Kranjc, Yu, & Hodgetts, 2004). Classrooms with poor acoustics are common (Knecht, Nelson, Whitelaw, & Feth, 2002) and rarely meet the guidelines set forth by the American Speech-Language-Hearing Association (2005) or American National Standards Institute (2002) for unoccupied noise levels, reverberation, or signal-to-noise ratios (SNR). Performance decrements in noisy classrooms are even more concerning for young children (i.e., < 5 years) who show significantly worse speech perception in noise than older children (Jamieson et al., 2004). In addition to classroom observations and teacher questionnaires, audiologists may use speech perception in noise tests to identify young and school-aged children who are at high risk for educational difficulties in noisy classrooms. As a result, the educational audiologist must have access to efficient, practical, portable, and sensitive speech-in-noise measures to quantify the behavioral effects of classroom noise. Valid and reliable quantification of listening difficulty often provides evidence of a child's need for instructional and communication accommodations, special education support, and hearing assistance technology (HAT) to enhance the SNR at the child's ear. Furthermore, sensitive speech-in-noise measures provide evidence to document benefits from HAT after it is fit on a child (American Academy of Audiology, 2008).

According to Elkins (1984) and Mendel and Danhauer (1997), sensitive speech perception tests are defined as those having the following characteristics: (1) a clear purpose, (2) identification of individuals for whom the test is designed, (3) evidence of validity, (4) confirmation of reliability through reports of typical variance and equivalent lists, and (5) defined procedures for administration, scoring, and interpretation. The validity and reliability of a test are particularly important because these data provide evidence that the test was constructed carefully and appropriately. Because most speech perception tests already typically address face validity (i.e., appears to be a good measure) and content validity (i.e., has appropriate content/stimuli), the most pertinent forms of validity to examine for this study include construct, convergent, discriminant, concurrent, and predictive validity. Construct validity confirms that a test measures what it is intended to measure. It may be examined through analyses of convergent validity (i.e., correlated to similar measures) or discriminant validity (i.e., not correlated to dissimilar measures). Concurrent validity examines whether a test will show significant differences in performance between groups of listeners that should be different, such as normal hearing and hearing impaired (Trochim, 2005). Although the aforementioned definition will be used for this critical review, concurrent validity may also be defined similarly to convergent validity, where results on the speech perception measure are compared to results on a similar assessment within the same testing period. Finally, predictive validity examines the relatedness of the test to a similar measure at a later time. For example, at a later testing period, the examiners could determine if the results on the speech perception test correlate to listening abilities in noise as reported on a subjective questionnaire.

Reliability, on the other hand, relates primarily to the repeatability of test results, equivalency among test items, and equivalency of lists in the test. Assessments of the latter two types of reliability are particularly important because items on a test, and lists within a test, must be equally intelligible (i.e., understandable) in background noise to allow for similar scores across lists or listening conditions (i.e., with and without HAT). In other words, the test must have inter-item and inter-list equivalence to have good test-retest reliability. Equal intelligibility in noise across test items and lists may not occur by simply equating for intensity or equal average root-mean-square (RMS) amplitude across the stimuli on a test (BKB-SIN, 2005; Nilsson, Soli, & Sullivan, 1994). As a result, most pediatric speech perception tests that are designed for use in quiet conditions do not have equivalent word lists when used with background noise. For example, the word lists for the Word Intelligibility by Picture Identification (WIPI; Ross & Lerman, 1970; Ross, Lerman, & Cienkowski, 2004) are not equivalent in broadband noise (Chermak, Wagner, & Bendel, 1988). Similarly, researchers found lack of list equivalence in noise for the Northwestern University-Children's Perception of Speech Test (NU-CHIPS; Chermak, Pederson, & Bendel, 1984; Elliott & Katz, 1980). Therefore, these tests, or any test that is not designed for use in noise, should not be used for assessing speech perception in noise as they may not allow for reliable comparisons of performance in various test conditions (such as aided and unaided) when using different lists.

Given the detrimental effects of noise on children's performance, the importance of quantifying speech perception in noise is clear. As a result, selection of sensitive speech perception tests is paramount for obtaining valid and reliable data. However, at this time, there are no cumulative peer-reviewed publications that critically analyze the construction and clinical utility of speech perception tests in noise for children. Therefore, the primary goal of this paper is to provide a critical review of speech perception measures in noise specifically designed for young and school-aged children. This review will present information regarding the sensitivity, validity, and reliability of available measures, as well as advantages and disadvantages of each test for examining pediatric speech perception in noise. Additionally, the clinical utility of two of the most sensitive measures for speech perception testing in the schools will be shown through case studies of three children who were assessed with and without frequency modulated (FM) systems.

Method

The speech perception tests in noise included in the critical review were identified through a comprehensive search of the literature using electronic databases (e.g., PubMed, ERIC) and a manual search of references or tests published from January 1970 through March 2010. Speech perception tests had to meet the following three criteria for inclusion: (1) design considerations for testing in noise, (2) stimuli with vocabulary levels appropriate

Table 1. Summary of Speech-Perception Tests in Noise for Children

Test (Acronym)	Ages	Test Description	Advantages	Disadvantages
Bamford-Kowal-Bench Speech in Noise test (BKB-SIN)	5+ years	Modified-adaptive test; Measures SNR loss for sentences in multi-talker babble	High validity, reliability, & sensitivity; may be used with any population; simple administration & scoring; portable & may be used in the classroom; inexpensive; on CD	May have ceiling/floor effects at standard SNRs; only appropriate for school-aged children
Hearing in Noise Test for Children (HINT-C)	6-12 years	Adaptive test; measures 50% correct threshold for sentences in speech-shaped noise	High validity, reliability, & sensitivity; computerized; may be used with any population; multiple languages; simple administration, scoring & interpretation	Expensive; only appropriate for school-aged children; speech-shaped noise may not be as challenging as other noises
Listening in Spatialized Noise-Sentences test (LiSN-S)	6-11 years	Adaptive test; measures sentence-in-noise thresholds for varying noise locations & types of noise	High validity, reliability, & sensitivity; computerized; simple administration, scoring, & interpretation	Only designed for use with suspected APD; may only present under headphones; expensive
Pediatric Speech Intelligibility test (PSI)	3-6 years	Measures percent-correct performance for words and sentences in single-talker competing noise	High validity & reliability, may be used with young children; simple scoring and interpretation; inexpensive; on CD	Complicated administration; may have ceiling/floor effects; single-talker noise may not be challenging; only for young children

Note. APD=auditory processing disorders; CD= compact disc; SNR=signal-to-noise ratio loss for children less than 12 years of age, and (3) availability for purchase. Once a test met the inclusion criteria, all publications related to the construction, validity, and reliability of that same test were identified. Each test was analyzed for its sensitivity using the recommendations by Elkins (1984) and Mendel and Danhauer (1997). Using these criteria, the following areas were addressed for each speech perception test in noise: (1) purpose and population, (2) validity and reliability, (3) administration, scoring, and interpretation, and (4) advantages and disadvantages of using the test in schools. In addition, three case studies were presented where sensitive measures were used to evaluate speech perception performance and potential benefit from FM systems.

Results

Commercially Available Speech-in-Noise Tests

As shown in Table 1, the literature review and manual search resulted in the identification of four speech perception tests in noise, including the Bamford-Kowal-Bench Speech-in-Noise (BKB-SIN) test, Hearing in Noise Test for Children (HINT-C), Listening and Spatialized Noise-Sentences test (LiSN-S), and the Pediatric Speech Intelligibility (PSI) test. Critical reviews for each of these tests will be provided in the following sections. In addition, brief reviews will be presented for three tests that were used in research studies, but are not commercially available for purchase. These tests include the Adaptive Spondee Test (AdSpon), Children's Realistic Index of Speech Perception (CRISP), and the Phrases in Noise Test (PINT).

Bamford-Kowal-Bench Speech-in-Noise Test (BKB-SIN)

Purpose and population. The purpose of the BKB-SIN test (BKB-SIN, 2005) is to determine the listener's signal-to-noise ration (SNR) loss, which is the increase in SNR that is required by a listener to obtain 50% of key words correct as compared to normative data from normal-hearing listeners of the same age (i.e., 5 to 6 years). In other words, using a formula to calculate SNR loss, this test determines the dB difference between a child's SNR for a 50% (SNR 50) correct level and the average SNR of children within a similar age range. The test consists of 18 list pairs (e.g., lists 1a and 1b are to be used together) of 10 sentences each spoken by a male speaker and in the presence of multi-talker babble. The stimuli are presented at pre-recorded SNRs that decrease 3-dB steps from a +21 to a -6 dB SNR. The BKB-SIN test may also be used to evaluate aided benefit, assess performance with directional microphones, and screen for auditory processing disorders. It was designed for children (≥ 5 years) or adults and for populations having normal-hearing, hearing loss (unaided), hearing aids, cochlear implants, and other auditory disorders (e.g., auditory processing deficits).

Validity and reliability. The sentences were determined originally from language samples of young children with hearing loss and are at a vocabulary level of a typical first-grade child (Bench & Bamford, 1979; Bench, Kowal & Bamford, 1979). Although construct validity was not addressed in the BKB-SIN user manual (BKB-SIN, 2005), it was assessed adequately in several publications. For example, convergent validity is shown in a study by Wilson, McArdle, and Smith (2007) who reported that scores on the BKB-SIN are within one standard deviation of scores on the HINT for adults with normal hearing and hearing loss. Discriminant validity was addressed, somewhat, in two studies that evaluated noise tolerance with a measure known as acceptable noise levels (ANL) and speech perception in noise using the BKB-SIN (Donaldson et al., 2009; Schafer & Wolfe, 2008). Both studies confirm that noise tolerance is not significantly correlated to speech perception on the BKB-SIN. These findings are also similar to what is reported for users of hearing aids (Nabelek, Freyaldenhoven, Tampas, Burchfield, & Muenchen, 2006). Concurrent validity is addressed in the user manual (BKB-SIN, 2005) by identifying significant performance differences between adults with normal hearing and cochlear implants, as well as among children with normal hearing in three age groups: 5 to 6 years, 7 to 10 years, and 11 to 14 years. Predictive validity is assessed in the Donaldson et al (2009) and Schafer and Wolfe (2008) studies with statistically significant correlations (i.e., correlation coefficients of 0.60 and 0.46, respectively) between performance on the BKB-SIN and subjective self-assessment questionnaires that measured ease of communication, speech recognition in reverberation, and social and emotional hearing handicap.

Test-retest reliability, as provided in the user manual (BKB-SIN, 2005) was high according to the results of testing 48 children with high levels of education and 44 children from lower-income families. In addition, the authors provided the estimated reliability based on the number of list pairs given. Because root-mean-square (RMS) equivalence of the sentences did not ensure equal intelligibility across the sentences, the creators grouped sentences with similar thresholds and grouped lists into pairs to ensure equivalent difficulty. The final BKB-SIN test provides equivalent list pairs that, according to normative data, do not deviate by more than 1 dB from the grand-average performance across lists.

Administration, scoring, and interpretation. Overall, the administration, scoring, and interpretation are presented clearly in the BKB-SIN manual (BKB-SIN, 2005). The procedures for administration of the test are the same as those used to collect the normative data. The scoring forms are easy to interpret, and the manual provides a chart to calculate the child's SNR loss (i.e., dB difference from children with normal-hearing). Although interpretation for children is not as clear as it is for adults, the creators suggest that SNR losses of children should be evaluated on a case-by-case basis along with supporting data, such as speech, language, and academic skills and learning environment.

Advantages and disadvantages. The BKB-SIN is a sensitive test that has data to support its validity and reliability in each of the critical areas. Because it is recorded on compact disc (CD), it may easily be used in the sound booth or for testing in the classroom using a portable CD player with detachable loudspeakers. There are two CDs provided from Etymotic Research with the stimuli on the same channel (Standard CD) or with the stimuli on separate channels (Split Track CD), which allows for testing with HAT and directional microphones on hearing aids. Another advantage of this test is the use of multi-talker babble, which is more difficult and realistic than most other types of background noise (Sperry, Wiley, & Chial, 1997). Overall, this is a well-constructed, flexible test for use with children.

One minor disadvantage of the BKB-SIN is that the listener could hit ceiling at the poorest SNR on the CD (-6 dB) or floor at the best SNR (+21 dB). However, the manual describes how the SNR can be adjusted to avoid this issue. This may be particularly relevant when testing the benefit of HAT, which can improve performance from the no-FM-system condition by 20 dB (Schafer & Thibodeau, 2006). In addition, the BKB-SIN may only be used for children who have receptive vocabulary levels of a typical five-year-old child, which further limits the appropriateness of this test to school-aged children.

Hearing in Noise Test for Children (HINT-C)

Purpose and population. The purpose of the HINT-C is to assess speech intelligibility and functional hearing of children, ages 6 to 12 years, in quiet and in speech-shaped noise, using an adaptive-testing paradigm to obtain a threshold at the 50% correct level. The HINT-C was developed using a subset of age-appropriate sentences in the HINT (Nilsson et al., 1994) that were separated into ten, ten-sentence lists with similar phonemic content (Nilsson, Soli, & Gelnett, 1996). Children are asked to correctly repeat the entire sentence. When testing in noise, the speech is fixed, typically at 65 dBA, and presented from a loudspeaker at 0 degrees azimuth. The noise, which is matched to the long-term-average spectrum of the sentences, is varied adaptively to find the child's threshold. Noise may be presented from the loudspeakers located at the front (0 degrees) or sides (90 or 270 degrees) of the child. The test was designed for use with any listener including those with normal hearing and hearing loss. The test manufacturer (Bio-Logic) also clearly states that the test may be used to asses speech intelligibility for children who are trying to learn in noisy classrooms, especially those who are English- language learners and those who have learning disabilities, otitis media, hearing aids, and/or cochlear implants.

Validity and reliability. In order to determine which of the sentences from the HINT were age appropriate, normal-hearing children ages 5 to 6 years were asked to repeat them in a quiet listening condition. If a child did not repeat the sentence correctly, it was discarded from the final version of the HINT-C. In terms of construct validity, convergent validity of the HINT-C was shown in the same study as discussed for the BKB-SIN (Wilson et al., 2007), while no evidence of discriminant validity was found. Concurrent validity was addressed for the HINT-C with comparisons between adults with normal hearing and children of different ages. Children, ages 6 to 12 years, showed significantly poorer performance than older children (> 13 years) and adults, and percentile rankings are provided for each age group in each listening condition (Nilsson et al., 1996). In addition, the HINT was shown to differentiate performance between 15 adults with normal hearing and nine adults with bilateral, symmetrical sensorineural hearing losses (Nilsson, Soli, & Sumida, 1995). Similar results are expected for the HINT-C when comparing performance of those with normal and impaired hearing. No direct evidence of predictive validity was found; however, listeners with cochlear implants, who had significantly poorer performance on HINT sentences in noise (fixed intensities) than those with normal hearing, reported significant difficulty listening in noisy situations via subjective questionnaires (Schafer & Thibodeau, 2004). Therefore, a relationship between speech perception performance in noise on the HINT and subjective, real-world difficulties likely exists.

Measures of reliability for the HINT-C are referenced back to the development of the original HINT (Nilsson et al., 1994). During the development of the HINT, test-retest reliability was confirmed by testing 18 adults with normal hearing (Nilsson et al., 1994) who showed average performance that only varied by 1 dB or less across lists. Similar findings were found in a later study (Nilsson et al., 1995). In addition, prior to this testing, the sentences and lists were equated for phonemic content, intelligibility, and difficulty.

Administration, scoring, and interpretation. Initially, the HINT-C was available as hardware and software, or on CD, where the examiner was required to adjust the signal levels manually using guidelines. However, the HINT was recently acquired by another manufacturer (Bio-Logic) and is now only available as a hardware and software system known as HINTPro. This system includes the HINT and HINT-C in 12 languages. When using the computerized format, administration, scoring, and interpretation is clear, understandable, and simple. The examiner is only asked to indicate if the child repeats the whole sentence correctly, and the software adjusts the sentence levels automatically to obtain the 50% correct threshold. All of the information needed for the interpretation of the person's threshold in noise is provided by the computerized program.

Advantages and disadvantages. Overall, the HINT-C has strong data to support its validity and reliability, as well as clear test administration, scoring, and interpretation. The HINTPro, which includes the HINT-C, is a flexible and portable system, which may be used under headphones or in the soundfield. In addition, it includes normative data for conditions with speech and noise from the same loudspeaker (0 degrees azimuth) and speech and noise from spatially separated loudspeakers (speech at 0 degrees and noise at 90 or 270 degrees azimuth). Soundfield testing may allow for assessment of aided benefit with hearing aids, cochlear implants, directional microphones, or HAT (spatially separated loudspeakers for the latter two). One unique aspect of HINTPro is the monitoring of patient reliability during testing. That is, if the child's responses are highly unreliable, testing will be automatically discontinued.

The main disadvantage to the HINTPro is the cost of the computerized system, which is approximately $5,000. This price may limit its use in school districts. Another disadvantage to the HINT-C is the use of speech-shaped noise, which is not as realistic or as challenging as multi-talker babble (Sperry et al., 1997). Finally, the test can only be used for children who have vocabulary levels greater than or equal to a typically-developing six-year-old.

Listening and Spatialized Noise-Sentences Test (LiSN-S)

Purpose and population. The North American LiSN-S is designed to assess a child's abilities to understand speech in the presence of noise arriving from different directions. The speech and noise stimuli are presented via headphones using a computerized program that creates the perception of a three-dimensional acoustic space. The program uses an adaptive-testing paradigm to determine if a child receives an advantage from spatially separated speech and noise sources. The four listening conditions tested include speech presented from the front (i.e., 0 degrees azimuth) and differing types of noise (i.e., noise from same or different voices) presented from varying locations. As the child repeats what he or she hears, the examiner records the number of correctly-repeated words into the program. The test was designed to assess children ranging from 6 to 11 years suspected of having auditory processing disorders. The test can also be used following some type of intervention to examine improvements in this area of binaural auditory processing in noise. Normative data for older children and adults are now available.

Validity and reliability. The sentences for the LiSN-S were written by Australian speech-language pathologists who specialized in the rehabilitation of children with hearing loss (Cameron & Dillon, 2007), and they were constructed according to procedures used to develop the original BKB sentences (Bamford & Wilson, 1979). Construct validity for this test is difficult to examine because it is a fairly new measure and the only adaptive test in noise designed specifically to assess auditory processing disorders. Two initial studies with some evidence regarding convergent and discriminant validity are currently available. An examination of convergent validity was attempted by comparing the normative data from the North American version of the LiSN-S to the normative data from the Australian version of the LiSN-S; however, unexplainable significant differences were found between the two groups of children (Cameron et al., 2009). Discriminant validity was determined in a study by Cameron and Dillon (2008) where children's results on the Australian version of LiSN-S were compared to four other common measures for assessing auditory processing disorders (i.e., Dichotic Digits, Masking Level Difference, Pitch Pattern Sequence, and Random Gap Detection Test). The researchers hypothesized that the LiSN-S examined different auditory processes than the other measures, which was confirmed by a lack of significant correlations (i.e., correlation coefficients ranged from 0.05 to 0.5). Concurrent validity was reported via significant differences across ages of typically developing, normal-hearing children (Cameron & Dillon, 2007; Cameron et al., 2009). In addition, significant performance differences were reported for children with suspected auditory processing disorders and those with no listening difficulties (Cameron & Dillon, 2008). Although no direct measure of predictive validity was found, the Cameron and Dillon (2008) study showed a relationship between performance on the LiSN-S and abnormal auditory behaviors from children with suspected auditory processing disorders.

According to Cameron and colleagues (2009), test-retest reliability of the LiSN-S was fairly high according to testing with 36 children with normal hearing and auditory processing. Correlation coefficients were significant for four of the five testing conditions and ranged from 0.5 to 0.7. List equivalency was also confirmed in this study with 24 children with normal-hearing sensitivity.

Administration, scoring, and interpretation. The administration of LiSN-S is fairly simple in that it only requires the examiner to enter the number of words repeated correctly within each sentence. During the adaptive testing, the noise remains constant (55 dB SPL), and the sentences are adapted to determine a speech reception threshold in noise after presenting 22 to 30 sentences. The testing takes approximately 20 minutes. The LiSN-S uses a unique scoring technique to reduce effects of language, learning, and communication abilities, which involves computing difference scores. These difference scores represent the spatial advantage (i.e., scores with noise at 0 degrees minus scores with noise at $\pm$ 90 degrees), talker advantage (i.e., scores with noise from same talker minus scores with the noise from different talkers), and total advantage. The interpretation of the scores is automated. The computerized program determines if the child's score is within or outside of the normal range when compared to normative data from North American children, and it creates a report (Cameron et al., 2009).

Advantages and disadvantages. The primary advantages of this test are the use of adaptive stimuli, which avoids ceiling and floor effects (i.e., 100% and 0%, respectively) and the computerized administration, scoring, and interpretation. Additionally, the test has strong validity and reliability and may be used for children with suspected auditory processing disorders before and after treatment or therapy. The use of headphones to present stimuli has several advantages when compared to use of loudspeakers because headphones eliminate variability associated with child head movement during testing, remove limits resulting from loudspeaker and listener placement issues in the soundfield, and reduce effects of reverberation.

The primary disadvantages to the LiSN-S is the limited population for which it was designed and the inability to present the test using loudspeakers. The test was not designed for children with hearing loss, hearing aids, or cochlear implants; yet, these populations of children exhibit great difficulty listening in noisy situations. While the use of headphones does reduce variability in several domains, the test would have greater application to other populations if normative data were provided in the soundfield using loudspeakers. Another disadvantage is the price of the program, which is approximately $1,000. Children with suspected auditory processing disorders are only one small group of children served by an educational audiologist. Therefore, the cost of the program may outweigh the benefits of having the test, especially when there are less expensive tests that can be used in the classroom.

Pediatric Speech Intelligibility Test (PSI)

Purpose and population. The purpose of the PSI is to examine diagnostic speech intelligibility of young children, ages 3 to 6 years, using a closed set of monosyllabic words and sentences in quiet and noise conditions (Jerger & Jerger, 1982, 1984). Children are asked to listen to the speech stimulus presented in quiet or in single-talker competing noise. They are then asked to indicate their responses by pointing to the corresponding noun (one of five pictures) or sentence (one of five pictures) depicted on a color picture card. Stimuli may be presented via headphones or one loudspeaker located at 20 cm and 90 degrees azimuth from the child. The examiner uses various fixed signal levels to obtain a performance-intensity function (i.e., performance at various percent-correct scores) in quiet or in noise for dichotic testing (headphones only). The test is designed for young child with normal hearing or hearing loss.

Validity and reliability. The vocabulary for the PSI was developed from language testing and samples of 87 children, ages 3 to 6 years. Construct validity, or more specifically the convergent validity, of this test was difficult to determine because there were no other speech perception tests in noise for children at the time of its development. However, a recent study used PSI to examine the effects of early amplification on speech perception performance in noise of young children with mild to profound hearing loss (Sininger, Grimes, & Christensen, 2010). Although the PSI was not correlated to any of the other measures in the study (i.e., no convergent validity; correlation coefficients ranging from 0.003 to 0.034), this finding was expected because the other measures focused on other aspects of speech perception, production, and language (i.e., speech perception of contrasts in quiet, speech production, and receptive and expressive language). The fact that the PSI was not correlated to these measures shows that it has discriminant validity. Concurrent validity was also confirmed in this recent study because the PSI was able to differentiate between children with good and poorer speech perception in noise. Concurrent validity was also shown via significantly different scores in noise across ages (Jerger & Jerger, 1984) and between children with normal-hearing sensitivity and otitis media (Jerger, Jerger, Alford, & Abrams, 1983). Predictive validity was revealed in the Sininger et al. (2010) study because the age at amplification was a significant predictor of PSI performance. In other words, the PSI was sensitive for identifying the expected effect of more positive outcomes for earlier amplified children.

The reliability of the measure was clearly addressed in the test manual (Jerger & Jerger, 1984). Test-retest reliability was confirmed with 35 children with normal hearing and 18 children with varying degrees of sensorineural hearing loss. Correlation coefficients were high for words and for sentences (i.e., .82 to .96). In addition, equivalence of word and sentence lists was established with children with normal hearing on two occasions. The test developers also examined practice effects with children with normal hearing and concluded that three practice trials would essentially eliminate any influence of practice effects on performance. Inter-item equivalency was also examined in noise, and the words and sentences were found to be equivalently difficult for children in the competing noise.

Administration, scoring, and interpretation. When compared to the previous three speech perception measures, the administration of the PSI is somewhat complicated. The user manual provides step-by-step instructions, but the rules for changing intensities when obtaining the performance-intensity functions are difficult to follow. However, the scoring (i.e., percent correct) and the interpretation are straightforward. Normative data are provided for children with normal hearing, and the interpretation of these data is fairly clear. The Sininger et al. (2010) paper used a modified approach to the PSI by presenting speech from a loudspeaker at 0 degrees azimuth and noise from a speaker at 180 degrees azimuth in quiet and at +10, 0, and -10 SNRs. This testing technique may be easier to administer clinically; however, the applicability of the normative data may be influenced by using different loudspeaker arrangements and conditions from the original design.

Advantages and disadvantages. The primary advantage to this test is that it may be used with younger children, ages 3 to 6 years. In addition, the use of closed-set materials and administration of the practice lists prior to testing addresses issues related to speech intelligibility (i.e., production) and vocabulary level. The original test was available on cassette tape, but it is now available on CD along with the manual and picture cards from Auditec (www.auditec.com). Finally, the PSI appears to be well-constructed, and there is adequate data to support its reliability and validity.

The administration of the test is not as simple as other tests, and the suggested loudspeaker arrangements in the manual may limit the applicability of the test for determining benefit from HAT and directional microphones. However, the equipment set-up described in the Sininger et al. (2010) study would address this issue. It is also possible that the use of a single competing talker may not adequately predict common listening situations where more than one talker is present. Finally, the use of fixed signal levels in some of the conditions may lead to ceiling and floor effects. This issue is addressed partially by always using a variety of SNRs for each child, if time permits. On the other hand, use of several SNRs would increase administration time and might lead to issues with attention span.

Tests Developed for Research Studies

As shown in Table 2, the literature review identified three additional tests that were developed for use in research studies, which are not currently available for purchase (i.e., not sold commercially). These include the Adaptive Spondee Discrimination (AdSpon) Test, Children's Realistic Intelligibility and Speech Perception (CRISP) Test, and the Phrases in Noise Test (PINT).

Adaptive Spondee Discrimination Test (AdSpon) and Children's Realistic Intelligibility and Speech Perception Test (CRISP)

The first two tests, the AdSpon (Galvin, Hughes, & Mok, 2010) and the CRISP (Litovsky, 2003, 2005), both used a computerized four-alternative, forced-choice paradigm with simple spondees in the presence of noise. In both of these tests, children were asked to indicate the spondee they heard on a computer screen, and both used an adaptive-testing technique to obtain a speech-in-noise threshold at the 79.4% correct level. The AdSpon spondees were presented in the presence of speech-shaped noise, and it is unknown whether the examiners equated the spondees for intelligibility or average RMS. The speech was presented from a loudspeaker at 0 degrees azimuth, while the noise was presented from a loudspeaker ± 90 degrees azimuth. Galvin and colleagues (2010) used the AdSpon stimuli with children older than 10 years. No validity or reliability data were discussed. The CRISP was used in conjunction with various types of noise for research purposes (Litovsky, 2005; Litovsky, Johnstone, & Godar, 2006), but the same phrases were used in each study and were equated for average RMS. Speech was presented from a loudspeaker at 0 degrees azimuth, while noise was presented from a loudspeaker to the front, right, or left side of the child. This test was designed for children 4 years and older. No reliability or validity data were found for this test either. One major concern about these two tests is related to the intelligibility of the phrases in the various background noises. As mentioned previously, equating for RMS does not ensure equal intelligibility in noise. Scaling procedures or adjustments must be made to ensure equal intelligibility in any type of background noise, especially when using adaptive procedures to find a threshold in noise.

Table 2. Summary of Research-Based Speech Perception in Noise Tests for Children

Test (Acronym)	Ages	Test Description
Adaptive Spondee Discrimination test (AdSpon)	≥ 10 years	Computerized adaptive test; measures 74.4% correct threshold for spondees in speech-shaped noise at ± 90 degrees azimuth; children selects spondee on computer screen
Children's Realistic Intelligibility and Speech Perception test (CRISP)	4+ years	Computerized adaptive test; measures 74.4% correct threshold for spondees in various types of noise at 0, +90, -90 degrees azimuth; children selects spondee on computer screen
Phrases in Noise Test (PINT)	3+ years	Modified-adaptive test; measures 50% correct threshold for simple phrases in classroom noise at 180 degrees azimuth; child repeats phrase or acts it out with a doll

Phrases in Noise Test (PINT)

The PINT consists of ten simple closed-set phrases (equal duration) about body parts (e.g., brush his teeth) with four-classroom noise that is equated to the long-term average RMS of the phrases (Schafer, 2005). It may be used with children as young as 3 years of age (Schafer & Thibodeau, 2006). The intensity of the phrases was determined carefully using intensity-scaling procedures similar to those used for the development of the HINT (Nilsson et al., 1996). The slightly revised version of the test consists of six different lists of 24 randomly-selected phrases that are presented at 57 dBA and classroom noise that automatically adapts in 3-dB steps from a -18 to a +12 dB SNR. The children are asked to repeat the phrases, as well as act them out with a doll and several related objects (e.g., comb his hair). Each list takes approximately three minutes and yields a 50% speech-in-noise threshold. Previous data support convergent validity of the PINT threshold to thresholds obtained using an adaptive-testing technique for a similar test (HINT; Nilsson et al., 1996). In addition, the test has concurrent validity because it detects substantial performance differences between children with normal hearing and those with cochlear implants (Schafer, 2005) and between conditions with and without FM systems (Schafer & Thibodeau, 2006). The PINT is currently being revised, and normative data are being collected from children with normal hearing and hearing loss.

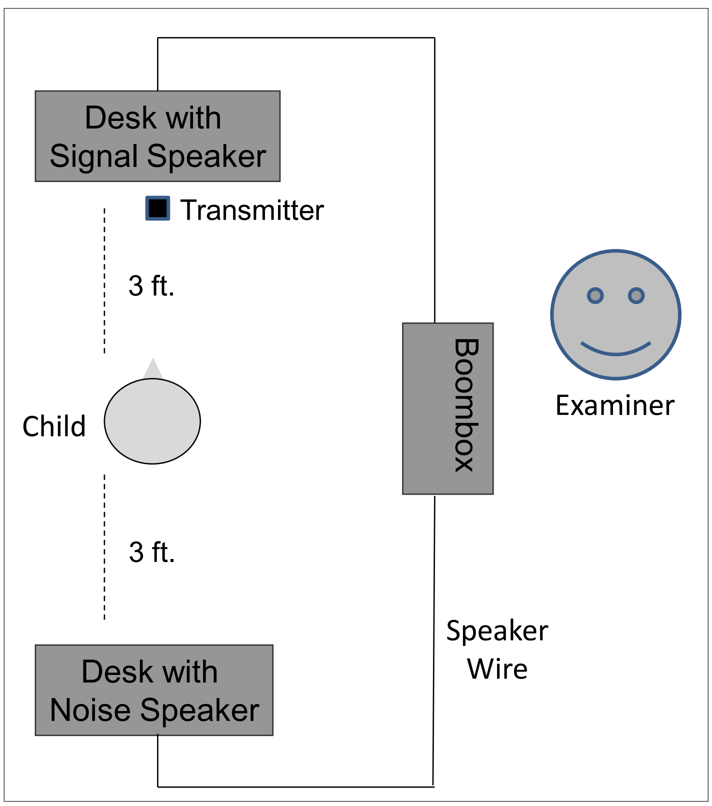

Figure 1. Simple equipment arrangement for conducting speech perception testing in noise in a classroom.

Case Studies

The first two case studies provide a summary of assistive technology evaluations to determine educational need for HAT in the classroom. These case studies provide representative examples of how speech perception measures in noise may be used as part of a functional evaluation for an FM system. Both evaluations included speech perception tests in noise with the BKB-SIN, classroom observations, and teacher questionnaires. As shown in Figure 1, the speech perception testing was conducted in the child's primary classroom using a simple soundfield arrangement. The BKB-SIN is presented via a portable CD player (e.g., Sony CFD-ZW755), two single-coned loudspeakers, and speaker wire to allow for a distance of three feet from the child's head. The levels of the loudspeakers are calibrated in dBA using an inexpensive sound level meter (e.g., Radio Shack Digital Display Sound Level Meter). When the FM system is in use, the transmitter lapel microphone is suspended six inches from the center of the single-coned loudspeaker. Boom or cheek transmitter microphones are placed three inches from the signal speaker. The examiner sits nearby to control the portable CD player and to record the child's responses on the scoring form. The third case study was conducted in a clinical environment to determine performance in noise and differences between two FM-system conditions.

Case One

Jim is a second-grade student who qualified for special education with learning disability and emotional disturbance eligibilities. An assistive technology referral was generated to evaluate his educational need for an FM system.

A hearing screening revealed normal-hearing sensitivity from 500 to 4000 Hz bilaterally. A functional evaluation was conducted and included teacher interviews, questionnaires, classroom observation, and speech perception in noise. Teacher interviews identified concerns about his inability to listen, focus, and participate in structured activities within the classroom. Teacher reports on the Screening Instrument for Targeting Educational Risk (S.I.F.T.E.R.; Anderson, 1989) questionnaire showed that he was at risk in the area of communication. During a classroom observation, Jim had some difficulty with directions and independently following along with the lesson. He required frequent one-on-one assistance from the teacher. In addition, he watched the actions of other students to follow class activities and transitions. He was seated with his back toward the whiteboard and teacher.

Jim's speech perception performance in noise was evaluated in his general education classroom using the BKB-SIN. His performance was assessed with and without a personal FM system (i.e., Phonic Ear Easy Listener and headphones) that was obtained from the school district's pool of back-up FM equipment. In the no-FM condition, the SNR associated with 50% of key words correct was +13.5 dB, which indicates a SNR loss of 12.7 dB. According to the BKB-SIN manual, children his age score an average of 0.8 dB (critical difference level of 3.5 dB), which suggests that he performs significantly worse than children his age when listening in noise. In addition, he will require an even better SNR than +13.5 to hear the full message from the teacher (i.e., greater than 50% of words) in a typical classroom. With the personal FM system set to a comfortable volume, Jim's performance at the 50% correct level improved to +7 dB. As a result, the FM system significantly improved his performance relative to no-FM performance; however, his performance is still well below average scores of peers in his age group.

Given the results of the speech perception testing in noise, teacher observation/questionnaires, and classroom observation, Jim has educational need for an FM system at school. A personal FM system (i.e., Phonak Edulink receiver) was recommended during direct instruction in his academic classes (i.e., math, reading, language arts). In addition, Jim would benefit from preferential seating in the classroom away from noise-producing equipment and with his body facing toward the whiteboard and teacher.

Case Two

Sam is a sixth grader with a bilateral mild-to-moderate sensorineural hearing loss in the right ear and moderate-to-moderately severe sensorineural hearing loss in the left ear. He wears his hearing aids during direct academic instruction, but not during his other classes. He receives some academic support (i.e., tutoring) in math, reading, and language arts. At the time of the evaluation, he did not qualify as a special education student with speech or auditory impairment. Special education teachers generated the assistive technology referral to determine if he had educational need for an FM system.

A functional evaluation was conducted and included teacher questionnaires, a classroom observation, and speech perception testing in noise. The S.I.F.T.E.R. questionnaires from his teachers indicated that he was at-risk in the areas of academics, attention, communication, and class participation when compared to peers. His teachers believed that an FM system could facilitate better attention and focus during class and ease his frustration. The observation revealed a non-carpeted classroom with hard-surfaced walls. One outside wall was near a busy road, and road noise was audible in the classroom. Sam sat at the front of the classroom during the lecture, but he moved to other desks when working collaboratively with peers. He required frequent direction to follow along with the teacher's lesson.

The BKB-SIN was used to measure Sam's speech perception in noise in his primary classroom. In a condition with his hearing aids, Sam required a +15.5 dB SNR to obtain 50% of key words correctly, while the average score for normal-hearing children his age is -0.9. Therefore, Sam's performance was significantly worse than normal-hearing children his age. When using a loaner FM system and his personal hearing aids (i.e., Phonic Ear Easy Listener with neckloop), his performance on the BKB-SIN improved to +10.5 dB, which indicated a significant improvement in performance with the FM system relative to the no-FM condition.

The results of the speech perception testing, teacher questionnaires, and classroom observations indicated educational need for an FM system during direct instruction. Other recommendations included continued preferential seating in the classroom and full-time use of the hearing aids.

Case Three

Sarah is a six-year-old child using a unilateral Advanced Bionics cochlear implant with an ear-level Auria sound processor. Her mother requested an appointment at the University of North Texas Speech and Hearing Center to determine if her school-issued personal soundfield FM system (i.e., Phonic Ear Toteable) was providing optimal benefit in noise. Sarah reported that she did not use her personal soundfield FM system consistently at school, especially during circle and center times in her Kindergarten classroom.

Speech perception testing in noise was conducted in the sound booth using the equipment arrangement shown in Figure 1. The PINT was used to determine Sarah's speech-in-noise threshold in three conditions: (1) cochlear implant alone, (2) personal soundfield FM system, and (3) personal FM system electrically coupled to the implant sound processor (i.e., Phonak MLx-S receiver and Campus S transmitter). Based on data published by Schafer and Thibodeau (2006), significant differences are indicated when there is a difference of 3.2 dB between two listening conditions.

In the cochlear-implant-alone condition, Sarah required a +10.5 dB SNR to repeat half of the phrases correctly. The personal soundfield FM system improved her threshold to +1.5 dB, while the personal FM system resulted in a threshold of -9 dB. Although the personal soundfield system significantly improved performance relative to the cochlear implant alone, the personal, electrically-coupled system had a clear and significant advantage over the soundfield system. Following the evaluation, a personal FM system was recommended for use at school. The personal system resulted in significantly better thresholds in noise than the personal soundfield system, and it provided Sarah more consistent access to the signal from the FM system when she was involved in listening to the teacher during centers or circle time.

Recommendations and Conclusions

The measurement of speech perception in noise provides individualized information regarding a child's ability to function in a noisy classroom. This measure may be used in conjunction with teacher questionnaires and classroom observations to provide a functional assessment in the child's customary learning environment. Specifically, speech perception testing in noise allows educational audiologists to quantify educational need for special education services and HAT (i.e., FM systems). Unfortunately, according to this critical review, there are only four published speech perception tests in noise that were designed for children and three additional tests that were used for pediatric research studies.

All of the published tests have high sensitivity, validity, and reliability, but they differed in presentation format, advantages, disadvantages, and clinical utility for schools. When purchasing a test for use in a clinical setting with a sound booth, the HINT-C has several advantages over other tests, including multiple languages and computerized administration and interpretation. However, it is an expensive test that utilizes speech-shaped noise. The cost of the test and use of a noise type that may not relate closely to the noise encountered in the classroom may limit its applicability for school-based audiology services. Similar to the HINT-C, the LiSN-S test has several advantages, including the ability to measure effects of type of noise and location of noise (i.e., 0 versus 90 degrees azimuth) and computerized administration, scoring, and interpretation. On the other hand, it was only designed for use with headphones and for children who have suspected auditory processing disorders. The PSI is a less expensive alternative to the aforementioned tests, but it may only be used with young children and has several disadvantages over other measures (Table 1), including ceiling and floor effects. Given these findings, the BKB-SIN appears to be the most viable choice for measuring speech perception in noise with school-aged children in the sound booth or in the classroom (Figure 1). It is fairly inexpensive and has straightforward administration and scoring. The possibility of ceiling and floor effects are addressed by using a modified-adaptive testing approach and an adjustable range of SNRs. As shown in these case studies, this test has been used successfully in the classroom to identify educational need and to examine benefits of FM systems for improving speech perception in noise. For younger children, the PSI is the only option at this time; however, normative data collection for the PINT is currently in progress. It is expected to be available for purchase within the next two years. In addition, it is possible that the other research-based tests, the AdSPON and CRISP, will also be available in the near future.

The results of this critical review highlight the significant need for additional sensitive tests to assess children's speech perception in noise. Specifically, there are very few tests that were designed for young children who have several special testing considerations (i.e., vocabulary, attention span, closed/open set). Although there are several published tests for young children (i.e., WIPI, NU-CHIPS), these tests were not designed for use in noise and, therefore, do not have equivalent word lists in the presence of background noise.

Although this critical review focused on speech perception in noise, which is at the identification level of Erber's hierarchal levels of auditory-skill development, it would also be beneficial for audiologists to have access to sensitive tests in noise at other levels including detection, discrimination, and comprehension (Erber, 1982). Assessment along the auditory-skill continuum will help determine educational need for children with a wide range of abilities and levels, as well as help to focus on auditory goals for the speech-language pathologist or educator. The only test commonly used to assess these hierarchal auditory skills in school-aged children - the Test of Auditory Comprehension (TAC; Trammel, 1981) - is no longer published, presents auditory stimuli via cassette tape, and may have outdated vocabulary, picture response options, and stimuli. In addition to these issues, the normative data from the 1980s are no longer applicable for children with newer digital hearing aids and cochlear implants who may have benefited from early hearing detection and intervention. As a result, future pediatric research should focus on the development of a hierarchical battery of tests for young and older school-aged children that includes speech perception measures in noise, as well as more sophisticated levels of auditory-skill development.

References

American Academy of Audiology Clinical Practice Guidelines. (2008, April). *Remote Microphone Hearing Assistance Technologies for Children and Youth from Birth to 21 Years*.

American National Standards Institute. (2002). *Acoustical performance criteria, design requirements, and guidelines for schools* (No. ANSI S12.60-2002). Melville, NY.

American Speech-Language-Hearing Association. (2005). *Position statement and guidelines for acoustics in educational settings.*

Anderson, K. L. (1989). *Screening Identification For Targeting Educational Risk (S.I.F.T.E.R.) in Children with Identified Hearing Loss.* Retrieved from http://www.kandersonaudconsulting.com/

Bamford, J., & Wilson, I. (1979). Methodological considerations and practical aspects of the BKB sentence lists. In J. Bench & J. Bamford (ESds.) *Speech-Hearing Tests and the Spoken Language of Hearing Impaired Children.* London: Academic Press, pp. 146-187.

Bench, J., & Bamford, J. (1979). *Speech-Hearing Tests and the Spoken Language of Hearing Impaired Children.* London: Academic Press.

Bench, J., Kowal, A., & Bamford, J. (1979). The BKB (Bamford-Kowal-Bench) sentence lists for partially-hearing children. *British Journal of Audiology, 13,* 108-112.

BKB-SIN: Bamford-Kowal-Bench Speech in Noise Test. (2005). Elk Grove, IL: Etymotic Research.

Bradley, J.S. & Sato, H. (2008). The intelligibility of speech in elementary school classrooms. *Journal of the Acoustical Society of America, 123(4),* 2078-2086.

Cameron, S., & Dillon, H. (2007). Development of the Listening in Spatialized Noise-Sentences Test (LiSN-S). *Ear and Hearing, 28(2),* 196-211.

Cameron, S., & Dillon, H. (2008). The Listening in Spatialized Noise-Sentences test (LiSN-S): Comparison to the prototype LISN and results from children with either a suspected (central) auditory processing disorder or a confirmed language disorder. *Journal of the American Academy of Audiology, 19(5),* 377-391.

Cameron, S., Brown, D., Keith, R., Martin, J., Watson, C., & Dillon, H. (2009). Development of the North American Listening in Spatialized Noise-Sentences test (NA LiSN-S): Sentence equivalence, normative data, and test-retest reliability studies. *Journal of the American Academy of Audiology, 20(2),* 128-146.

Chermak, G.D., Pederson, C.M., & Bendel, R.B. (1984). Equivalent forms and split-half reliability of the NU-CHIPS administered in noise. *Journal of Speech and Hearing Disorders, 49(2),* 196-201.

Chermak, G.D., Wagner, D.P., & Bendel, R.B. (1988). Interlist equivalence of the word intelligibility by picture identification test administered in broad-band noise. *Audiology, 27(6),* 324-333.

Donaldson, G.S., Chisolm, T.H., Blasco, G.P. Shinnick, L.J., Ketter, K.J., & Krause, J.C. (2009). BKB-SIN and ANL predict perceived communication ability in cochlear implant users. *Ear and Hearing, 30(4),* 401-410.

Elkins, E. (Ed.). (1984). *Speech recognition by the hearing-impaired.* ASHA Reports*, 14,* 2-15. Rockville, MD: ASHA.

Elliot, L.L., & Katz, D.R. (1980). *Northwestern University Children's Perception of Speech (NU-CHIPS).* St Louis, MI: Auditec.

Erber, N. (1982). *Auditory training.* Washington, DC: Alexander Graham Bell Association for the Deaf.

Galvin, K.L., Hughes, K.C., & Mok, M. (2010). Can adolescents and young adults with prelingual hearing loss benefit from a second, sequential cochlear implant?. *International Journal of Audiology, 49(5),* 368-377.

Jamieson, D.G., Kranjc, G., Yu, K., & Hodgetts, W.E. (2004). Speech intelligibility of young school-aged children in the presence of real-life classroom noise. *Journal of the American Academy of Audiology, 15(7),* 508-517.

Jerger, J., & Jerger, S. (1984). *Pediatric Speech Intelligibility Test*. St. Louis: Auditec.

Jerger, S., & Jerger, J. (1982). Pediatric speech intelligibility test: performance-intensity characteristics. *Ear and Hearing, 3*(6), 325-334.

Jerger, S., Jerger, J., Alford, B. R., & Abrams, S. (1983). Development of speech intelligibility in children with recurrent otitis media. *Ear and Hearing, 4(3),* 138-145.

Knecht, H.A., Nelson, P.B., Whitelaw, G.M., & Feth, L.L. (2002). Background noise levels and reverberation times in unoccupied classrooms: Predictions and measurements. *American Journal of Audiology, 11(2),* 65-71.

Litovsky, R.Y. (2003). Method and system for rapid and reliable testing of speech intelligibility in children. U.S. Patent No. 6,584,440.

Litovsky, R.Y. (2005). Speech intelligibility and spatial release from masking in young children. *Journal of the Acoustical Society of America, 117(5),* 3091-3099.

Litovsky, R.Y., Johnstone, P.M., & Godar, S.P. (2006). Benefits of bilateral cochlear implants and/or hearing aids in children. *International Journal of Audiology, 45(Suppl1),* S78-S91.

Mendel, L.L. & Danhauer, J.L. (Eds.) (1997). *Audiologic Evaluation and Management and Speech Perception Assessment.* San Diego, CA: Singular Publishing Company, Inc.

Nabelek, A.K., Freyaldenhoven, M.C., Tampas, J.W., Burchfield, S.B., & Muenchen, R.A. (2006). Acceptable noise level as a predictor of hearing aid use. *Journal of the American Academy of Audiology, 17(9),* 626-639.

Nilsson, M., Soli, S.D., & Sullivan, J.A. (1994). Development of the Hearing in Noise Test for the measurement of speech reception thresholds in quiet and noise. *Journal of the Acoustical Society of America, 95(2),* 1085-1099.

Nilsson, M.J., Soli, S.D., & Sumida, A. (1995). Development of norms and percent intelligibility functions for the HINT. *House Ear Institute.*

Nilsson, M.J., Soli, S.D., & Gelnett, D.J. (1996). Development of the sharing in noise test for children (HINT-C). *House Ear Institute.*

Ross, M., & Lerman, J. (1970). A picture identification test for hearing-impaired children. *Journal of Speech and Hearing Research, 13,* 44-53.

Ross, M., Lerman, J., & Cienkowski, K.M. (2004). *Word Intelligibility by Picture Identification – WIPI, 2nd Edition.* St. Louis, MO: Auditec.

Schafer, E.C., & Thibodeau, L.M. (2004). Speech recognition abilities of adults using cochlear implants interfaced with FM systems. *Journal of the American Academy of Audiology, 15(10),* 678-691.

Schafer, E. C. (2005). Improving speech recognition in noise of children with cochlear implants: Contributions of binaural input and FM systems. *Dissertation Abstracts International, 66*(2), 789 (UMI No. ATT 3163263).

Schafer, E.C., & Thibodeau, L.M. (2006). Speech recognition in noise in children with bilateral cochlear implants while listening in bilateral, bimodal input, and FM-system arrangements. *American Journal of Audiology, 15,* 114-126.

Schafer, E.C., & Wolfe, J. (2008). Acceptable noise levels in adults with cochlear implants. *Canadian Hearing Report, 3(2),* 32-35.

Sininger, Y.S., Grimes, A., & Christensen, E. (2010). Auditory development in early amplified children: Factors influencing auditory-based communication outcomes in children with hearing loss. *Ear and Hearing, 31(2),* 166-185.

Sperry, J.L., Wiley, T.L., & Chial, M.R. (1997). Word recognition performance in various background competitors. *Journal of the American Academy of Audiology, 8(2),* 71-80.

Trammell, J. L. (1981). Test of Auditory Comprehension. North Hollywood, CA: Foreworks.

Trochim, W.M. (2005). *Research Methods: The Concise Knowledge Base.* Cincinnati, OH: Atomic Dog Publishers.

Wilson, R.H., McArdle, R.A., & Smith, S.L. (2007). 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, 50(4),* 844-856.