

ABSTRACT

Title of Document: RELIABILITY AND TEST ENVIRONMENT
OF THE SCAN-A WITH CHILDREN AGES
12-15

Michele Lynn Spencer, Doctor of Clinical
Audiology, 2007

Directed By: Professor Sandra Gordon-Salant, Hearing and
Speech Sciences

The SCAN-A is a widely utilized auditory processing screening tool for use with adolescents and adults 12-to-50 years of age. The SCAN-A consists of four subtests: Filtered Words, Auditory Figure-Ground, Competing Words, and Competing Sentences, and takes about 20 minutes to administer. Other versions of this screening tool exist (e.g., SCAN and SCAN-C) that are standardized for use with children under 12 years of age. However, previous reports indicate that test-retest reliability is poor and test environment affects performance by young children. In this study, the effect of test environment (sound attenuating booth versus quiet room) and test-retest reliability for the 12-to-15 year old age group was investigated. Thirty participants, ages 12-to-15 years old, who were normally developing, were tested using the SCAN-A four times, twice in both a quiet room and a sound attenuating booth, with testing in both environments conducted one month apart. A high false-positive rate (43% of participants) was found for the first administration of the SCAN-A, with fewer

participants identified with possible APD with subsequent test administrations. Results revealed a significant main effect of test administration time, and no significant main effect of test environment or significant interaction, for the Filtered Words, Auditory Figure-Ground, Competing Words, and Total Test standard score. No significant main effects or interaction was found for the Competing Sentences subtest. This investigation demonstrates that the SCAN-A has low specificity, a high false-positive rate, and poor test-retest reliability.

RELIABILITY AND TEST ENVIRONMENT OF THE SCAN-A WITH
CHILDREN AGES 12-15

By

Michele Lynn Spencer

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park, in partial fulfillment
of the requirements for the degree of
Doctor of Clinical Audiology
2007

Advisory Committee:
Professor Sandra Gordon-Salant, Chair
Professor Tracy Fitzgerald
Professor Rochelle Newman
Professor Froma Roth
Professor Deborah Speece

© Copyright by
Michele Lynn Spencer
2007

Table of Contents

Table of Contents	ii
List of Abbreviations.....	iii
Chapter 1: Introduction.....	1
Chapter 2: Literature Review.....	7
Identifying an Auditory Processing Disorder with Screening Measures.....	6
Test Measures Developed to Identify CANS Lesions.....	9
What is an Auditory Processing Disorder Test Actually Measuring?.....	12
Diagnosing an Auditory Processing Disorder using a Test Battery Approach.....	13
SCAN: A Screening Test for Auditory Processing Disorders.....	22
Other investigations using the SCAN.....	26
SCAN-C: A Test for Auditory Processing Disorders in Children.....	32
SCAN-A: Test for Auditory Processing Disorders in Adolescents and Adults.....	36
Summary.....	39
Chapter 3: Research Questions	41
Chapter 4: Method	44
Participants.....	44
Stimuli.....	45
Procedure.....	46
Preliminary Testing.....	46
Experimental Procedures.....	47
Scoring Procedures.....	49
Chapter 5: Results.....	51
Chapter 6: Discussion.....	67
Effect of Test Administration.....	67
Effect of Test Environment.....	70
Total Test Standard Score by Administration.....	72
SCAN-A Outcomes.....	75
Limitations of the Current Study.....	78
Conclusions.....	78
Appendices.....	80
References.....	85

List of Abbreviations

1. APD: Auditory Processing Disorder
2. BF1: Binaural Fusion subtest number one
3. BF2: Binaural Fusion subtest number two
4. BFT: Binaural Fusion Test
5. CLD: Classroom learning disability
6. DDT: Dichotic Digits Test
7. FST: Filtered Speech Test
8. LD/LD: Children with learning disabilities believed to have an auditory dysfunction
9. LD/N: Children with learning disabilities with normal auditory function
10. LPFS: Low Pass Filtered Speech
11. MLD: Masking Level Difference
12. NL: Normal learning ability
13. N/N: Children who are normally developing with normal auditory function
14. NU-6: Northwestern University Auditory Test Number Six
15. NU-CHIPS: Northwestern University Children's Perception of Speech
16. OM: Otitis Media
17. PPT: Pitch Pattern test
18. PPVT-R: Peabody Picture Vocabulary Test- Revised
19. RASP: Rapidly Alternating Speech Perception
20. ROC: Receiver operating curve
21. SCAN: A Screening Test for Auditory Processing Disorders

22. SCAN-A: Test of Auditory Processing Disorders in Children
23. SCAN-C: Test of Auditory Processing Disorders in Adolescents and Adults
24. SSW: Staggered Spondaic Word Test
25. TC: Time-compressed speech
26. WISC-R: Weschler Intelligence Scale for Children

Reliability and Test Environment of the SCAN-A with
Children Ages 12-15

Adolescents with possible auditory processing difficulties are a highly misunderstood population in the field of audiology. This is in large part because of the lack of consensus regarding the exact definition of an auditory processing disorder (APD). In 1996, the American-Speech-Language-Hearing Association Task Force on Central Auditory Processing defined the disorder as a deficit in one or more of the following behaviors (p. 41):

- Sound localization and lateralization
- Auditory discrimination
- Auditory pattern recognition
- Temporal aspects of audition, including
 - temporal resolution
 - temporal masking
 - temporal integration
 - temporal order
- Auditory performance decrements with competing acoustic signals
- Auditory performance decrements with degraded acoustic signals

In April 2000, the American Academy of Audiology held an auditory processing consensus conference to evaluate APDs in school-aged children and at that time defined an APD as:

A deficit in the processing of information that is specific to the auditory modality. The problem may be exacerbated in unfavorable acoustic environments. It may be associated with difficulties in listening, speech understanding, language development, and learning. In its pure form, however, it is conceptualized as a deficit in the processing of auditory input. (Jerger & Musiek, 2000, p. 468)

The beginning of auditory processing testing began in the 1950s when Bocca and colleagues (Bocca, Calero, & Cassinari, 1954; Bocca, Calero, Cassinari, Migliavacca, 1955) developed a filtered speech test to discern the auditory function of patients with central lesions. The interest in testing complex auditory abilities was first mentioned by Myklebust (1954). Myklebust (1954) indicated there was a need for developing tests that evaluate individuals with possible communication disorders.

Other investigators interested in auditory processing testing followed, and many tests of auditory processing abilities were developed using participants with documented lesions. These tests included Dichotic Digits Test (Kimura, 1961), Competing Sentences Test (Willeford, 1977), Dichotic Consonant-Vowel Test (Berlin, Cullen, Hughes, Berlin, Lowe-Bell, & Thompson, 1975), Synthetic Sentence Index Test (Jerger & Jerger, 1974, 1975), and the Frequency (Pitch) Pattern Sequence Test (Pinheiro & Ptacek, 1971; Ptacek & Pinheiro, 1971). The aim of these tests was to identify the side and site of central nervous system lesions (Goldstein, 1961; Goldstein, Goodman, & King, 1956; Jerger, 1960) and intracranial lesions (Calero & Antonelli, 1968; Korsan-Bengsten, 1973; Lynn & Gilroy, 1972; Lynn & Gilroy, 1977; Musiek, Wilson, & Pinheiro, 1979), because routine audiometric tests were not sensitive to these lesion sites.

Accuracy of a test in diagnosing a disorder is determined by the test's sensitivity and specificity rates. Sensitivity rates, the ability of a test to identify accurately an individual with APD, and specificity rates, the ability of the test measure to identify accurately an individual without APD, have been reported for certain APD test measures. For example, the reported sensitivity and specificity rates

for the Dichotic Digits test are 75% and 85% (Musiek, Gollegly, Kibbe, & Verkest-Lenz, 1991), respectively. The sensitivity has been reported as 25% and the specificity has been reported as 100% for the Competing Sentences subtest (Domitz & Schow, 2000).

The test measures that were developed to diagnose the side and site of central nervous system lesions were subsequently applied to identify APDs in children. Evaluating children and adolescents with a “possible central auditory processing disorder” using central auditory testing was first mentioned at the symposium on Central Auditory Dysfunction (Keith, 1977). However, this application of tests to diagnose auditory system lesions may not be entirely appropriate for children who have no known pathology (lesions) of the central auditory nervous system. It was assumed that if children exhibit poor performance on these types of measures, it reflects a problem in the neurological “wiring” or neural transmission of auditory signals.

Auditory processing disorders (APD) can affect people of any age; however, diagnosing APDs in school-aged adolescents is important because APDs affect an individual’s listening and learning abilities, thus potentially negatively influencing an individual’s performance in school (Amos & Humes, 1998; Chermak, 1996). Teachers and parents routinely describe an individual with an APD as having difficulty with one or more of the following: (a) understanding speech when background noise is present; (b) following verbal directions; (c) determining exactly what they hear; and (d) understanding speech in situations where speech information is degraded (Jerger & Musiek, 2000).

A multidisciplinary approach is recommended to determine an individual's auditory processing ability. This team of experts should include speech-language pathologists, psychologists, and audiologists. However, the final diagnosis of an APD should only be made by an audiologist (Bellis, 1997). For example, a speech-language pathologist may examine written and oral language ability in an individual suspected as having an APD. The speech-language pathologist may diagnosis the individual as having an APD by these tests measures; however, tests of written and oral language are not tests designed to diagnose APDs (Bellis, 1997). Nevertheless, it is common for a diagnosis of APD to result in delivery of speech-language pathology services for these children.

Currently, few measures are available to screen individuals for APD. The most widely utilized screening measures are the Test of Auditory Processing Disorders in Children (SCAN-C, Keith, 2000) and the Test of Auditory Processing Disorders in Adolescents and Adults (SCAN-A, Keith, 1995). The SCAN-C has four subtests: Filtered Words, Auditory Figure-Ground, Competing Words, and Competing Sentences. Each SCAN-A subtest was included in the test battery to assess a particular auditory process. These processes include auditory closure and binaural separation. The Filtered Words subtest is an auditory closure task, wherein the stimuli are low-pass filtered at 750 Hz, thus reducing the ability to clearly hear the test items. The Auditory Figure-Ground subtest is a measure in which the test items are presented in background noise. This is considered an auditory closure task, as well as a selective attention task (Clark, 1980). The Competing Words subtest is a dichotic listening task. The participant is asked to repeat two different words, one

presented to each ear. Finally, the Competing Sentences subtest is a measure of binaural separation. The individual hears two sentences, one in each ear, and is asked to repeat only one sentence.

According to the manual, the SCAN-C can be administered to children between 5-to- 11-years of age. The SCAN-A is comprised of the same subtests and is administered to adolescents and adults, ranging in age from 12-to-50-years (Keith, 1995). Keith (1986) developed the original SCAN for administration to children between 3-to-11- years of age. However, as a result of poor test-retest reliability (Amos & Humes, 1998) and effects of test environment on scores (Emerson, Crandall, Seikel, & Chermak, 1997), this version was revised by Keith (2000) and called SCAN-C. Other screening measures also have been proposed such as the Selective Auditory Attention Task (Cherry, 1980).

Although the SCAN-A is a widely utilized screening tool, only one study has assessed its validity and reliability (Keith, 1995). However, this investigation failed to include test-retest reliability data for the 12-to-18-year old age group. As stated previously, concerns also have been cited by Emerson et al. (1997) regarding the effect of test environment (sound attenuating booth versus quiet room) on the SCAN. Neither of these concerns has been addressed with the SCAN-A. The SCAN and SCAN-A both include Filtered Words, Auditory Figure-Ground, and Competing Words subtests. The test items for the SCAN and SCAN-A are the same. The present study addresses these concerns by: (a) determining test-retest reliability with adolescents ages 12-to-15; and (b) determining if there is a significant effect of test

environment (sound-attenuating booth versus quiet room) for adolescents between 12-to-15-years of age.

Literature Review

Identifying an APD with Screening Measures

Because of the significant amount of time involved in testing an individual using a test battery approach, it is important to develop screening measures prior to the full evaluation to quickly determine who is at risk for an APD. Adequate screening measures must have a high sensitivity rate and a high specificity rate. Sensitivity is the ability of a test to identify accurately an individual with an APD. Specificity is defined as the ability of the test measure to identify accurately an individual without an APD. Tests with low sensitivity will miss affected individuals, whereas tests with low specificity identify many unaffected individuals with a possible disorder, resulting in a high false positive rate and subsequent over-referrals. The test-retest reliability of a screening measure is also important to monitor changes over time.

Screening for APD is important, especially in the school-aged population. Auditory processing disorders can be diagnosed in young children (5-to-11 years of age), as well as adolescents (12-to-17 years of age). An APD can substantially affect a child in school or at home. Children identified with an APD can be easily distracted, have difficulty in background noise situations, have difficulty concentrating, and/or perform below grade level in school (Chermak & Musiek, 1992; Emerson, Crandall, Seikel, and Chermak, 1997). Thus, screening measures for APD are helpful to identify children with a possible APD before an extensive test battery is performed.

A screening test quickly identifies individuals who are at risk for the disorder and who may require evaluation or monitoring. According to Musiek et al. (1990),

the following conditions should be evaluated to determine if a screening measure for a particular measure is necessary: (a) the impact of the disorder on the individual; (b) the prevalence of the disorder in the overall population; (c) the ability to differentiate individuals with the disorder from individuals without the disorder; (d) the availability of diagnosis and treatment for the disorder; and (e) the benefit of the entire process, including screening, diagnosis, and treatment, should exceed the overall cost expended. It is estimated that 2-3% of children are identified with an APD (Chermak & Musiek, 1997), and this prevalence may be higher because some children are undiagnosed with a possible APD. Current treatments for children with APD include preferential seating and FM systems in the classroom. The report of the Consensus Conference on the Diagnosis of APD in School-aged Children (Jerger & Musiek, 2000) also mentioned screening measures for APD in children, suggesting that the screening measure should address hearing and/or understanding when background noise is present, comprehending degraded speech, understanding verbal instructions, differentiating and distinguishing sounds of speech, and providing unpredictable answers to stimuli.

The consensus committee recommended six conditions that must be considered when developing a screening measure for APD (Jerger & Musiek, 2000). First, the tasks should assess complex auditory stimuli using tests of both temporal and spatial resolution. Second, appropriate psychometric standards should be followed, including sensitivity and specificity, interobserver reliability, intertest consistency, and validity. Third, certain variables, including the number of items, presentation level, and type of response should be addressed. Fourth, factors that can

confound screening results, including training examiners, the presence of conductive and/or sensorineural hearing loss, and test environment, should be evaluated. Fifth, the linguistic, cognitive, and attentional demands needed for the screening measure should be minimal to limit confounding of test results. Finally, the screening measure should be brief, approximately 8-to-12-minutes in length. Additionally, a strong test-retest reliability coefficient is required when developing screening measures for APD (Cacace & McFarland, 1998).

Reliability can be defined as the consistency, dependability, stability, reproducibility or predictability of test scores (Cordes, 1994) with repeated test administrations. This definition of reliability indicates that the results will be repeated if the participant was tested again. The “observed” score on a test in relation to the “true” score on a test, is another definition of reliability. An “observed” score is the score that the patient received and the “true” score is the score that the patient should have received. A “true” score will be the same regardless of test environment, test administrators, and other factors; whereas the “observed” score may be biased by test administrators, test environments, and so on. An observed score is considered reliable if the variance of the observed score is within the variance of the true score (Cordes, 1994).

There are four types of reliability. These types are test-retest, parallel forms, inter-rater, and internal consistency reliability. Test-retest reliability is a measure of how stable scores are from one test administration to the next. Parallel forms of reliability are concerned with how equivalent one measure is with another. Inter-rater reliability is defined as how consistent scores are between raters of a test measure.

Lastly, internal consistency reliability determines how dependable the test items are across a test measure.

The validity of a test cannot be determined without reliable results (Cacace & McFarland, 1995). Validity is defined as the ability of the test measure to accurately identify the disorder it proposes to identify. The test measure should have high sensitivity, which means it should correctly identify individuals with the disorder, and high specificity, which means the test measure should correctly identify individuals without the disorder. Although several screening measures for APDs have been developed, none meets the criteria outlined above.

Test Measures Developed to Identify Central Auditory Nervous System Lesions

Central auditory testing was first described by Bocca, Calero, and Cassinari in 1954. They utilized a filtered speech test to assess the auditory function of individuals with temporal lobe lesions. The phonetically balanced, disyllabic words were low-pass filtered at 800 Hz. Results of this preliminary study revealed scores that were consistently lower bilaterally for individuals with intact central auditory nervous systems than for individuals with temporal lobe lesions. For most participants with temporal lobe lesions, discrimination was poorer than normal in the ear contralateral to the lesion and better in the ear ipsilateral to the lesion.

Multiple investigators have researched filtered word testing since Bocca et al. (1954). This research has indicated that low-pass filtered speech is sensitive to both brainstem (Calero & Antonelli, 1968; Lynn & Gilroy, 1977) and cortical lesions (Bocca, 1958; Hodgson, 1967; Korsan-Bengtson, 1973); therefore, the filtered speech test is unable to localize lesions of the central auditory system, but is helpful in

identifying the presence of lesions that may exist. The filtered word test assesses an individual's auditory closure. Auditory closure is the ability to understand speech without the entire speech spectrum available (Bellis & Burke, 1996) and requires the individual to attempt to determine what information is missing from the item; therefore, auditory closure tasks may involve receptive language processing (Keith, 1994).

Auditory Figure-Ground testing assesses an individual's ability to understand speech in the presence of a background competitor. Some investigators consider this another measure of auditory closure, because masking of the target speech signal by a noise background reduces the availability of speech cues; however, the individual must also have the ability to identify phonemes and words (Chermak & Musiek, 1997; Rintelman, 1985). Other investigators suggest that a speech-in-noise test assesses the ability of a listener to attend selectively to a target message and ignore irrelevant information (Clark, 1980). This view of the Auditory Figure-Ground test suggests that it has a cognitive component and is not auditory specific. Similar to filtered words, auditory figure-ground testing cannot determine site of lesion, but is helpful in determining if a lesion is present in the central auditory system from the eighth nerve to the temporal lobe (Olsen, Noffsinger, & Kurdziel, 1975).

Competing word testing assesses an individual's dichotic listening ability (Musiek, 1999). The participant is asked to repeat two different words, one presented to each ear (Musiek & Pinherio, 1985). Dichotic listening tasks have been reported to be sensitive to cortical lesions (Goodglass, 1967; Musiek, 1983a; Oxbury & Oxbury, 1969). Studies have reported that when a lesion is present, poorer results are found

for the ear contralateral to the lesion (Kimura, 1961) than in the ear ipsilateral to the lesion. This is because the contralateral pathway is more dominant than the ipsilateral pathway (Rosenweig, 1951) during dichotic listening tasks (Musiek & Pinherio, 1985). The sensitivity and specificity for the Dichotic Digits Test has been reported as 75% and 85%, respectively (Musiek, Gollegly, Kibbe, & Verkest-Lenz, 1991). Keith (1986) reported that the Competing Words subtest helps to assess an individual's neuromaturation, which is defined as the functional development of the central nervous system.

Competing sentences assess an individual's binaural separation. Binaural separation, also known as directed listening, is defined as the ability of an individual to hear and repeat what is heard in one ear only, and is considered another type of dichotic listening task (Musiek, 1999). Other researchers (Moray, 1959; Treisman & Geffen, 1968) have suggested that the Competing Sentences subtest is also a measure of selective attention and listening. The specificity and the sensitivity of the Competing Sentences subtest for individuals with APD has been reported as 100% and 25%, respectively (Domitz & Schow, 2000). The reliability of the competing sentences subtest in identifying lesions is questionable when compared to other dichotic tasks (Lynn & Gilroy, 1972; Musiek, 1983a).

The use of testing other modalities (e.g., language, vision, tactile) is also important when testing for APD. The next section will describe this in further detail.

What is an APD Test Actually Measuring? Is Modality Specificity Important in Diagnosing an APD?

Measures of auditory processing are intended to assess perception of auditory signals, and tap into deficits in the auditory pathways. However, McFarland and Cacace (1995) argue that an APD test battery should include tests that assess other factors, such as language, vision, and attention, to determine if the individual has a deficit in any other area that could contribute to their poor performance on an APD test battery. In their opinion, a test should be modality- specific.

Because of a lack of modality specificity in current APD tests, deficits in cognition, language, and attention cannot be ruled out (Cacace & McFarland, 1998). Poor performance on one measure within a test battery is often used to diagnose an individual with an APD. It is possible that poor performance on a single test could be associated with some non-auditory problem, such as a cognitive disorder.

McFarland and Cacace (1995) mention three categories of individuals who could possibly be diagnosed with an APD: (a) individuals who have a modality specific auditory perceptual deficit; (b) individuals who have both an auditory perceptual deficit, as well as other processing problems (e.g., vision); and (c) individuals who have a supramodal deficit, defined as a problem with motivation, attention, and/or memory that manifests as poor performance on an APD test battery. Most studies investigating APD test measures only assess an individual's abilities in the auditory modality and do not take into account other factors. Therefore, it is impossible to determine if an APD truly exists until other factors are ruled out.

Another issue with current APD tests is that these measures were first developed to assess individuals with known lesions of the central auditory pathway (Speaks, 1980). Subsequently, these same measures were utilized with individuals suspected of having an APD without documented lesions (Musiek, Baran, & Pinheiro, 1990). The underlying mechanisms for poor performance on APD tests in these individuals are unknown.

Another complicating factor is that comorbidity may exist, where an APD can be present with other disorders in areas such as language and attention (Musiek, Bellis, & Chermak, 2005). However, recent studies (Poremba et al., 2003; Salvi et al., 2002) have found that auditory tasks, such as listening in noise, actually stimulate the language, attention, and memory areas of the brain. Therefore, determining the contribution of language, attention and memory in auditory processing tasks may be more difficult than previously assumed.

Diagnosing an APD Using a Test Battery Approach

Developing and selecting the most appropriate test battery to diagnose an adolescent or child with an APD is a difficult but important task. Despite the variety of opinions on the nature of APDs, investigators all agree that a test battery approach is important to accurately identify an APD. Researchers in the field of audiology have tried to determine the best test battery for APDs. However, no gold standard exists, and there are many opinions on an appropriate test battery for identifying an APD. Following is a review of six articles that discuss the varying opinions on the most appropriate test battery for APD.

Musiek, Geurkink, and Kietel (1982) researched seven popular auditory processing measures [Rapidly Alternating Speech Perception (RASP), Binaural Fusion, Low Pass Filtered Speech (LPFS), Staggered Spondaic Word Test (SSW), Competing Sentences, Dichotic Digits, and Frequency Patterns] in an attempt to determine an appropriate test battery that would identify the largest number of children with APD. Using this test battery, they assessed 22 participants with normal peripheral hearing (12 boys and 10 girls ranging in age from 8-to-10-years) who were chosen because of difficulty in school. The researchers determined that a score for either ear that was at least one standard deviation below the standard mean of established normative values would result in a failure for that individual test. The normative values were determined for the 8, 9, and 10- year-olds independently. Results revealed that 86.4% failed the Competing Sentences test, 72.7% failed the Frequency Patterns test, 63.6% failed the Dichotic Digits test, and 50% failed the SSW test. These four tests were deemed the most sensitive in detecting children with possible auditory processing difficulties by the researchers. The LPFS, Binaural Fusion, and the RASP were the least sensitive with a failure rate of 18.2% for each measure. Only one participant passed all measures on the test battery.

The authors concluded that the most appropriate test battery would include the following measures: Competing Sentences, Frequency Patterns, Dichotic Digits, and SSW. They cautioned, however, that a test battery for APDs should not be developed based on this information alone because of the small sample size, as well as a lack of a “gold standard” test for auditory processing. Because there is no gold standard auditory processing test, determining false positives for this test battery is difficult.

The results described above could be typical for all children who have trouble in school, regardless of an APD. It is difficult to conclude from this study whether or not this is an acceptable test battery for APD on these results alone.

The small age group tested also is a limiting factor. These results can only be generalized to children 8-to-10 years of age. Another issue with this study is the minimal description of the procedure utilized. There is no mention in the procedures section that the test conditions were randomized across participants. Failure to randomize or counterbalance could result in outcome scores that are unreliable because of a learning effect or a participant's fatigue. The researchers concluded that four tests are the most sensitive in determining an APD; however, the children were not diagnosed with an APD, but were experiencing difficulty in school. It can be assumed that all children who have difficulty in school do not have an APD. Children can have difficulty in school for a number of reasons, which include attention deficit hyperactivity disorder, autism, dyslexia, as well as other learning disabilities.

A study by Ferre and Wilber (1986) investigated a test battery approach for identifying "auditory perceptual dysfunction" in children with learning disabilities utilizing auditory processing tests. Participants were 39 children 8-to-12- years of age, with normal hearing (pure tone thresholds less than 20 dB from 250-4000 Hz), normal middle ear function, no air-bone gaps greater than 5 dB, and an excellent speech recognition score, defined as 90-100%, on the Northwestern University Children's Perception of Speech (NU-CHIPS; Elliott & Katz, 1979) presented at 40 dB SL re: speech reception threshold. The children were assigned evenly into three

groups: normally developing with normal auditory function (N/N) (13 participants), learning disabilities with normal auditory function (LD/N) (13 participants), and learning disabilities believed to have an auditory dysfunction (LD/LD) (13 participants) and were matched for age, sex, race, and socioeconomic status. The children were previously diagnosed with learning disabilities using the Illinois Test of Psycholinguistic Abilities (Kirk, McCarthy, & Kirk, 1968) or the Detroit Test of Learning Aptitude (Baker & Leland, 1968) and the Weschler Intelligence Scale for Children (WISC-R; Weschler, 1974).

The investigators determined a child's auditory function by using a battery of measures which included: Peabody Picture Vocabulary Test-Revised (PPVT-R; Dunn & Dunn, 1981), Token Test for Children (DiSimoni, 1978), Goldman-Fristoe-Woodcock Auditory Sequential Memory Test (Goldman, Fristoe, & Woodcock, 1976), and the Goldman-Fristoe-Woodcock Auditory Discrimination in Quiet subtest (Goldman, Fristoe & Woodcock, 1976). This test battery was chosen to determine if the child had an auditory deficit in single-word comprehension (PPVT-R), sentence-lead connected language comprehension (Token Test for Children), retention of words in a series (Goldman-Fristoe-Woodcock Auditory Sequential Memory Test), and/or word discrimination (Goldman-Fristoe-Woodcock Auditory Discrimination in Quiet Subtest). Children with learning disabilities who scored one standard deviation below the mean on two or more of these tests were placed in the auditory dysfunction group. Each of these auditory tests was presented to each participant on the first day of testing in a randomized order.

The battery of tests was administered to the participants approximately one-to-two weeks after the first session. The test battery consisted of the lists from the NU-CHIPS adjusted to create the following subtests: (a) a low pass filtered speech test (LPFS) with a cut-off frequency of 1000 Hz and a 48 dB/octave rejection rate; (b) monotic time-compressed speech test using 60% time-compression; (c) dichotically presented binaural fusion test with a low-band-pass portion (BF1; frequencies between 350 and 650 Hz) and a high-band-pass portion (BF2; frequencies between 1850 and 2150 Hz) of the subtest; and (d) dichotic monosyllables test. All testing was performed by one experimenter. The order of test items was randomized for each participant.

Results of the investigation revealed a significant main effect for group (LD/LD, LD/N, N/N) and test condition (subtests). There were no significant interactions. Post-hoc testing revealed that performance for the LD/LD group was significantly different than the N/N group for all subtests. No statistically significant differences were found between the LD/N group and the N/N group, and between the LD/N and LD/LD group. Post-hoc testing, using Newman-Keuls, showed that all participants exhibited poorer performance on the dichotic monosyllables subtest compared to any other subtest.

The most sensitive subtest of auditory perceptual dysfunction was the LPFS test, followed by the BF2 subtest, the time-compressed subtest, and the dichotic monosyllables test. Most of the LD/LD participants (92.3%) performed below the norms on the LPFS subtest. Results of the BF1 subtest indicated that the LD/LD group performed below normal limits. For the LPFS, BF1, BF2, and the dichotic

monosyllables subtest, the LD/N group performed below normal limits. An equal number of participants from both the LD/LD and LD/N groups performed below normal limits on the time-compressed speech subtest.

Results revealed that each of the participants in the LD/LD group failed at least one subtest, with 92.3% of the participants failing more than one subtest. Eighty-five percent of the LD/N group failed one subtest, and 46.2% failed more than one subtest.

The investigators concluded that LPFS, BF, and time-compressed speech test were the easiest subtests for all participants, with all participants experiencing the most difficulty with the dichotic monosyllables test. They hypothesized that participants may have this difficulty because recognizing competing stimuli could be a higher-level auditory processing ability that may not be developed fully by 8-to-12-years of age.

The researchers also mentioned that the scores from the LD/LD group and N/N group overlapped, as did the scores from the LD/N and N/N groups. However, they stated that if the LD/LD group's performance is compared across all subtests, then the LD/LD group consistently falls below one standard deviation from the mean. Based on this evaluation, the investigators concluded that a single test is insufficient to diagnose possible auditory perceptual dysfunction.

The investigators also evaluated the false-positive rate on each subtest. They found that 62% of the LD/LD participants were identified correctly with the time-compressed speech test, and 92% of the LD/LD participants were identified correctly with the LPFS test. Although these two tests had relatively high hit rates for the

LD/LD group, the time-compressed speech test also incorrectly identified 62% of the LD/N participants (false positive), and the LPFS test incorrectly identified 23% of the LD/N participants (false positive). No false-positive results were mentioned for the N/N group. The researchers concluded that the time-compressed speech test is not as clinically useful as the LPFS test because the LPFS test had a higher hit rate for the LD/LD children and a lower false positive rate for the LD/N participants than the time-compressed speech test. The researchers also mentioned the pass/fail rate of the test battery. Only 4 of 13 participants failed all subtests. Therefore, they concluded that the best criteria to determine auditory perceptual dysfunction would be a failure on three or more subtests.

Although this study helped to determine a test battery to identify possible APDs, many problems with this study are evident. One is the use of only the NU-CHIPS as the speech material for all of the subtests. Because the same words were used to develop all the subtests, the children may have become familiar with the words. Another issue is the small sample size of each group ($n = 13$). With small groups, even one or two participants who fail or pass a certain test can skew the results. For every participant who passed one of the subtests, the passing percentage increased by 7%. A third issue is that the battery was administered at two test sites. Administering all of the subtests at both sites potentially is a fatal flaw in this investigation. If a study is administered at two different sites, there is a possibility that one of the sites is different than the other. If this is the case, the results obtained at one site can differ from those obtained at the other site. The investigators did not compare the results obtained at the two sites to verify that they were equivalent.

Singer, Hurley, and Preece (1998) designed a study to evaluate tests that would differentiate children with APDs from those without APDs, as well as investigate the possibility of combining measures to develop a test battery that would be more effective in differentiating these two groups. The study consisted of 238 participants, 7-to-13-years of age, who were assigned into two groups: (a) normal learning ability group ($n= 91$) and (b) a classroom learning disability group (CLD) ($n= 147$). Each participant in the classroom learning disability group was “presumed to have auditory processing problems (p. 74).” To participate in the study, each participant was required to have normal intelligence (did not specify how normal intelligence was defined), normal hearing (air conduction thresholds 15 dB HL or better from 250-8000 Hz), normal middle ear function, and normal word recognition ability (did not specify how normal word recognition was defined). Children were excluded from participation if they had a possible attention deficit problem or were bilingual.

Participants in the CLD group had difficulty with reading, following spoken directions, and maintaining attention in class. These participants were identified as having a possible APD by reports from teachers, speech-language pathologists, and psychologists.

The investigators administered seven tests to each participant: (a) the Binaural Fusion test (BFT, Willeford, 1977); (b) the speech recognition Masking Level Difference (MLD, Cullen & Thompson, 1974); (c) the Filtered Speech test (FST, Willeford, 1977); (d) time-compressed speech (TC) at 60% time compression (Beasley & Maki, 1976); (e) the Dichotic Digits Test (DDT, Musiek, 1983b); (f) the

Staggered Spondaic Word test (SSW); and (g) the Pitch Pattern test (PPT, Pinherio, 1977). All measures, except the BFT, were administered at 50 dB SL re: speech reception threshold. The BFT was presented at 30 dB SL re: speech reception threshold. The order of all tests was randomized between participants.

The investigators measured sensitivity using a variety of clinical decision analysis methods (Jerger, 1983; Schultz, 1972). The investigators performed analyses of hit rate and false positive rate to determine an appropriate test battery for APDs. The BFT, FST, and MLD were found to be the best measures to differentiate groups. Therefore, these three measures were selected as an appropriate test battery for APD. Based on cost-effectiveness results, Singer et al. (1998) recommended a test battery consisting of the BFT and MLD subtests in which a listener must fail both tests to be identified with an auditory processing disorder.

Numerous task forces and committees (American Speech-Language-Hearing Association, 1996; Jerger & Musiek, 2000; Musiek & Chermak, 1994) have met to develop position statements detailing the “correct approach” for measuring APDs in individuals. These position statements are not necessarily based on factual information, but rather are the opinions of the committee members. The committees and task forces all suggest the use of auditory, as well as electroacoustic and electrophysiological measures to identify an individual with an APD. The auditory measures recommended include pure tone threshold tests, speech recognition tests, and tests to determine auditory processing ability (e.g., localization and lateralization tests, time-compressed speech tests, competing speech tests, filtered speech tests). Auditory brainstem response, middle latency response, immittance and otoacoustic

emissions are suggested as electrophysiologic test measures, that should be obtained during an APD evaluation. No research studies have been developed to investigate the clinical usefulness and validity of the recommended test batteries developed by the Task Force on Central Auditory Processing Consensus Development (American Speech-Language-Hearing Association, 1996) or the Consensus Conference on the Diagnosis of Auditory Processing Disorders in School-aged Children (Jerger & Musiek, 2000).

A test battery involves a lengthy evaluation procedure. To avoid extensive testing on individuals who may not be at risk for APD, audiologists have developed screening techniques to quickly identify children and adults who have a high probability of having an APD. One popular screening test is the SCAN (Keith, 1986) and its derivatives, the SCAN-C (Keith, 2000) and the SCAN-A (Keith, 1995).

SCAN: A Screening Test for Auditory Processing Disorders

The SCAN: A Screening Test for Auditory Processing Disorders was developed by Keith (1986) as a screening tool for use in a school setting to determine possible APDs in children 3-to-11-years of age. The SCAN is fast, efficient, and takes approximately 20 minutes to administer in any quiet room using a cassette player.

The SCAN audiocassette consists of three subtests: (a) Filtered Words; (b) Auditory Figure Ground; and (c) Competing Words. A five-second silence interval occurs between each test item to ensure ample time for a participant's response. A 10-sec calibration tone of 1000 Hz also is included on the audiocassette.

The first subtest, Filtered Words, is comprised of 40 monosyllabic words filtered using a 1000 Hz low-pass filter with a roll-off of 32 dB/octave. The next subtest, Auditory Figure Ground (AFG), consists of 40 monosyllabic words that are presented with an ipsilateral background of mulitalker speech babble fixed at a +8 dB signal-to-babble ratio. The last subtest, Competing Words, is comprised of 50 monosyllabic word pairs presented simultaneously to each ear. There are two conditions for the Competing Words subtest. During the first condition, the child is asked to repeat the word they heard in their right ear first followed by the word they heard in their left ear. The second condition involves a left-ear first task or a directed left ear task, where the child is asked to repeat the word they heard in their left ear first followed by the word they heard in their right ear. For each subtest the child must repeat the entire word correctly to receive credit for that test item; however, for the Competing Words subtest, the child receives credit for each word repeated correctly regardless if only one word is repeated or if the words are repeated in the incorrect order.

The speech stimuli for the SCAN were taken from the Spache readability word lists (Spache, 1953) and the Kindergarten Phonetically Balanced word lists (Haskins, 1949). Each of these lists consists of monosyllabic words. Some of the words are used in more than one subtest of the SCAN.

The standardization of the SCAN (Keith, 1986) involved 1,034 normally achieving children (490 males, 535 females, and 9 gender not reported) with normal hearing and normal middle ear function from 3-to-11-years of age. After the standardization was complete, norms for the SCAN composite raw score and subtest

scores were developed. Separate norms were computed for each age between 3 years, 0 months to 9 years, 11 months, 30 days. Norms for the age range 10 years, 0 months to 11 years, 11 months, 30 days also were computed. However, the normative data for the 3 years, 0 months to the 4 years, 11 months, 30 days should be used cautiously because many of the children had difficulty completing tasks (Keith, 1986).

A subsequent study investigated the test-retest reliability of the SCAN (Keith, 1986). Test-retest reliability data were analyzed with 68 children in first and third grade only with a test-retest interval of approximately 6 months. Test-retest reliability coefficients for first grade students were .42, .40, .73, and .65 for Filtered Words, Auditory Figure-Ground, Competing Words, and the Composite Score, respectively. The test-retest reliability coefficients for the third grade students were .22, .41, .44, and .42 for the Filtered Words, Auditory Figure-Ground, Competing Words and Composite score, respectively. Each of these test-retest correlation coefficients, except for the third grade Filtered Words coefficient, was significant at either the .05 or .01 alpha level. However, a high correlation ($r > 0.8$) is desirable during test-retest analysis (Amos & Humes, 1998). According to Keith (1986), the test-retest interval may have been too long, thus giving the children time to mature and increase their scores from the first to the second test administration.

To determine if gender differences existed between males and females on the SCAN, 384 pairs of males and females (768 individuals) were matched based on test site, race, and age. Gender effects were significant for some age groups on all subtests: (a) Competing Words (ages six, seven, and eight); (b) Filtered Words (age

11); (c) Auditory Figure Ground (age 11); and (d) Composite Score (age seven, age 11). Females consistently performed higher than males on these subtests. Even though gender differences were significant, Keith (1986) concluded that gender does not need to be considered when interpreting SCAN scores because the significant effects were not consistent across subtests and age groups. The gender effect on the composite score for the 11-year-old-age group was significant at the $p < .01$ level. All other results were significant at the $p < .05$ level. Although the results were not consistent across subtests and age groups, it appears that there are significant gender differences on this test, suggesting that separate norms should be developed for boys and girls.

To test the construct validity of the SCAN, Keith (1986) conducted an analysis to determine the age-to-age progression of scores for each subtest and the composite. Significant Pearson product-moment correlations were found between age and subtest score, as well as between age and the composite score of the SCAN. The results therefore indicated that the raw scores on the SCAN increase from one age to the next.

Clearly, there are many problems with the original SCAN test. One concern is that some of the words used for the Competing Words subtest also were used for the Filtered Words and Auditory Figure-Ground subtests. Using the same words multiple times in a 20 minute test may make the test too easy because of learning effects. Alternatively, the use of the same word repeatedly can bias the results if the child is unfamiliar with the vocabulary item.

Another concern is that test-retest reliability data were obtained for only first- and third-grade students, rather than for all age-levels. Moreover, the test-retest reliability for the first- and third-graders was poor. It is critical to know whether or not test-retest reliability is comparable for all age groups for whom the test was developed.

Other investigations using the SCAN. Since the SCAN was developed in 1986, other studies have evaluated reliability issues with the SCAN. Emerson, Crandall, Seikel, and Chermak (1997) utilized the SCAN in two studies. One experiment assessed the difference in performance between children with otitis media (OM) and children with possible APDs when using the SCAN as the diagnostic tool. The second experiment examined whether a significant difference existed in subtest and composite scores measured from children in two different test environments: a sound-attenuating booth versus a quiet room.

The first experiment evaluated 28 children with normal hearing. The SCAN and the Peabody Picture Vocabulary Test-Revised (PPVT-R, Dunn & Dunn, 1981) were administered to gender and age-matched participants. The PPVT-R was administered with the SCAN to attempt to differentiate between an APD and a language problem. A child was considered to be in the OM group if, before the age of three, he or she experienced six or more episodes of OM. A child was considered to be in the non-OM group if he or she experienced two or fewer episodes of OM before the age of three. The test order consisted of the SCAN and then the PPVT-R for each participant. Results for the first experiment indicated that the SCAN composite score was not significantly different between the OM group and the non-

OM groups. However, there was a significant difference in performance of the groups between the Filtered Words and Auditory Figure-Ground subtests, as well as between the Competing Words and Auditory Figure-Ground subtests, with the OM group performing more poorly on the measures than the non-OM group. The results also indicated that six participants in the OM group and four participants of the non-OM group exhibited composite scores that fell two standard deviations below the norm for their age. As a result, all of these children were classified in the “disordered range.” Results with the PPVT-R revealed a significant difference between the OM group and the non-OM group. The researchers concluded that the PPVT-R differentiates the two groups but the SCAN does not.

The data from this study are limited in several ways. The researchers did not assess hearing sensitivity at 500 Hz in their hearing screening protocol, nor did they screen using tympanometry. If a child had a conductive hearing loss it would be unknown at the time of testing. This hearing loss could reduce the presentation intensity level of the test items, making the stimuli more difficult to hear. Another limitation is the number of OM episodes between the non-OM and OM groups. The differences in OM episodes could have been slight (two episodes for the non-OM group and three episodes for the OM group) and this could result in the lack of significant performance differences between the two groups on the SCAN. Also, the researchers did not randomize the order of tests conditions in this study. Finally, the PPVT-R does not definitively measure a language disorder (Altepeter, 1989).

Experiment two examined the effects of two testing environments, a quiet room and a sound-attenuated booth, on the subtest and composite scores of the

SCAN. This experiment utilized six children with normal hearing, five girls and one boy, ages 5.9 to 11.8 years of age. All children were clients of a speech-language pathologist and were tested in both a sound-attenuating booth and a quiet room with an ambient noise level of 28-34 dBA. APD status of the participants was not reported. The room administration order was randomized to avoid order-effects. A one-week testing interval was used between administrations. Although the group size of this experiment was small, the results revealed lower (poorer) scores on the SCAN administered in the quiet room than in the sound attenuating booth.

Specifically, five of the six children had lower scores on the Auditory Figure-Ground subtest and the SCAN composite score when tested in the quiet room. Two children were identified as having a possible APD from the composite score of the SCAN when tested in the quiet room; however, no children fell within the disordered range when they were tested in the audiometric booth. Unfortunately, statistical analyses were not performed because of the small subject sample. Another possible flaw of this study was that the children had a speech or language disorder, which could affect their verbal responses on the test. Therefore, their scores may not reflect their true auditory processing abilities.

Amos and Humes (1998) investigated the test-retest reliability of the SCAN with first- and third-graders. The purpose of this study was to retest the same two age groups evaluated by Keith (1986), but with a smaller test-retest interval. The test-retest interval chosen for this study was six to seven weeks. The investigators hypothesized that this interval would limit all unwanted effects, including learning and maturation, which was a concern in the study performed by Keith (1986).

Twenty-five first-graders and 22 third-graders who were Caucasian, spoke English, had normal hearing (screened at 20 dB for the frequencies of 1000, 2000, and 4000 Hz), and were performing at an academically age-appropriate level were administered the SCAN. All testing was performed in a quiet setting with ambient noise levels between 64 to 65 dBC and 50 dBA or less. The experimenter, time of testing, and the presentation level of 70 dB SPL were held constant across all participants. Results revealed that the participants' scores were higher on the second administration of the test for the Filtered Words subtest, Competing Words subtest, and the composite score. The results also revealed that no subtest or composite score of the SCAN produced a minimally acceptable test-retest correlation of 0.8. Correlation coefficients for the Filtered Words and Auditory Figure Ground raw and standard scores for both grades ($r < 0.35$) were not significant. Moderately strong and significant test-retest correlation coefficients for the Competing Words and composite score ($0.7 \leq r \leq 0.78$) were found for both grades. Higher test-retest correlation coefficients were found for both the Competing Words Subtest and the composite score.

Factor analysis also showed that the SCAN subtests are all related to the same underlying factor. The authors identified this factor as general auditory processing ability or speech-understanding ability; therefore, the three subtests are not distinguishable from one another.

Amos and Humes (1998) concluded that in order to measure a child's optimal performance, the SCAN must be administered twice, and the score on the second administration should be used to determine a child's estimated auditory processing

ability. Although the researchers did not test children with APD, they concluded that using the SCAN to evaluate the effectiveness of treatment of children with APD is very questionable because of the low test-retest correlation coefficients. Another issue with this study is that the researchers did not randomize subtest or test item order.

Smith, Bradham, Chandler, and Wells (2000) conducted a study to determine if there was a significant change in the subtest score of the SCAN when African American children were tested by both an African American and Anglo American examiner. The study sample included 47 African American children with normal hearing and normal development from the middle socioeconomic class (5-to-10-years of age). These children were assigned to one of three groups: (a) First grade level (children in kindergarten and first grade); (b) Second grade level (children in second and third grade); and (c) Third grade level (children in fourth and fifth grade). The SCAN was administered twice in a quiet room with a 15-minute test-retest interval. There were four pairs of examiners who were females (eight examiners total). Each pair consisted of one African American and one Anglo American examiner. The order of the examiner was randomized for each participant.

No significant effect of examiner race was found; however, participants performed significantly better on the second administration of the SCAN regardless of age. Results also indicated a significant learning effect for all age groups from first- to-second administration on all subtests and composite scores on the SCAN. There were no significant interactions. This result is consistent with results of Amos and Humes (1998). Nevertheless, some problems exist with this study. Hearing

thresholds were not tested for 500 Hz, and a tympanometric screen was not performed. The investigators do not mention methods of calibration nor the intensity level used in this study. There was no randomization of subtest order during this investigation. All of the children were middle class and do not represent the entire African American demographic that may be administered this measure.

Marriage, King, Briggs, and Lutman (2001) evaluated the reliability of the SCAN to determine if normative subtest and composite scores are valid with children who speak British English. One hundred and thirty-three children, ages 6 to 11, with normal hearing (defined as less than or equal to 25 dB HL from 500-4000 Hz), and normal middle ear function, were administered the SCAN in a quiet room. The researchers did not report any information regarding the children's performance and abilities in school or at home.

Results revealed that for the 9, 10, and 11 year olds, subtest scores and the composite score were significantly lower compared to the norms reported by Keith (1986), with the exception of the Competing Words subtest scores for the 11 year-old participants. For the Auditory Figure Ground subtest, significant differences were discovered for the six, seven, and eight year-olds when compared to the norms developed by Keith (1986). Significant differences, when using the normative data developed by Keith (1986), also were discovered for the Filtered Words subtest with the seven and eight year-olds. There were no significant effects of gender or subtest with the SCAN for any ages investigated. Results also revealed that word errors were present for each age group, with more word errors for the younger age groups. However, the word-error data were not analyzed statistically.

The investigators concluded from these results that published normative data for the SCAN cannot be used for British-speaking children. Accent and vocabulary errors contributed significantly to the poorer scores on the SCAN for British-speaking children than the normative sample of children tested in the United States.

There were many problems with the implementation of this study. First, the participants included children who were not native British English speakers whose data were not analyzed separately. Methodological problems included a variable presentation level and lack of randomization of subtest order.

SCAN-C: Test for Auditory Processing Disorders in Children

As a result of the Emerson et al. (1997) and Amos and Humes (1998) studies, Keith (2000) developed the SCAN-C: Test for Auditory Processing Disorders in Children (SCAN-C). There were many changes with the SCAN-C including: (a) rewording the test instructions; (b) recording to a compact disc (CD); (c) revising the Competing Words and adding the Competing Sentences subtest using the test items from the SCAN-A (Keith, 1995); and (d) obtaining normative data for children 5 years to 11 years, 11 months (instead of 3 years to 11 years, 11 months as in the original the SCAN).

Thus, the SCAN-C comprises the following four subtests: (a) Filtered Words; (b) Auditory Figure-Ground; (c) Competing Words; and (d) Competing Sentences. The Filtered Words and Auditory Figure-Ground subtest items are the same as those in the original SCAN, which were shown to have poor test-retest reliability with first- and third-grade children (Amos & Humes, 1998). The Competing Words subtest is similar to the original SCAN, but only 15 word pairs are presented to each ear

compared to 25 word pairs presented to each ear in the original SCAN. The Competing Sentences subtest consists of 10 sentence pairs that are presented dichotically. The child is asked to repeat the sentence heard in the target ear only. The stimuli utilized for the Competing Words and Competing Sentences subtest are taken from the SCAN-A (Keith, 1995). The Bamford-Kowal-Bench Standard Sentence Lists (Bench & Bamford, 1979) were used as test items for the Competing Sentences subtest.

Standardization studies of the SCAN-C included 650 children between 5 years, 0 months to 11 years, 11 months, tested by 114 examiners. Each child was tested in one of two testing environments, sound attenuating booth or quiet room, at the preference of the examiner. Participants were included in the study if they understood English, had normal hearing, and had intelligible speech. Middle ear function was not evaluated. Additionally, the children were not exclusively native speakers of English. Children were not excluded if they received special education services at school. This population was deemed representative of the general population (Keith, 1995).

As with the SCAN, normative values based on age were calculated for each subtest separately and for the composite score. Values were calculated for each age increment from 5 years, 0 months to 9 years, 11 months. Children 10 years, 0 months and children 11 years, 11 months were grouped into one normative category.

A test-retest reliability study was performed by testing 145 children between 5 to 11 years of age who were included in the original standardization sample. The participants were divided into two age groups depending on age, with one group

between five to seven years of age, and the other between 8 to 11 years of age. The median test-retest interval was 6.5 days, but ranged from two days to six weeks. The results of the test-retest reliability study revealed subtest and composite correlation scores ranging from .65 to .82 for the participants aged five to seven, and .67 to .78 for the participants aged 8 to 11 years. Most of these test-retest correlations are below the recommended .8 correlation coefficient (Amos & Humes, 1998). The order of subtest administration was not randomized.

The effect of test environment (sound-attenuating booth versus quiet room), was determined by obtaining scores from 54 children from a normative study (Keith, 2000) and placing them into two matched groups of 27 participants each based on age, gender, race/ethnicity, and parent education level. The children were assigned to two groups based on the environment in which they took the test (sound attenuating booth or quiet room). No significant differences in scores obtained in the two test environments were found for the SCAN-C.

Woods, Pena, and Martin (2004) researched possible sociocultural bias with the SCAN-C when comparing performances of Latino American children to Anglo American children. This study tested 20 Latino American and 20 Anglo American normally-developing children, with no known speech and/or language disorders, with normal hearing (hearing screening at 20 dB HL at 500 Hz, 1000 Hz, 2000 Hz, and 4000 Hz bilaterally) who were in second- and third- grade. The children were matched by socioeconomic status (SES). During testing, the participants were seated in a quiet room and the entire SCAN-C test was presented once at a comfortable

listening level. The listening level was not fixed, but was determined by the participant.

Results revealed no significant main effects or interactions of ethnicity and socioeconomic status (SES) on the composite score of the SCAN-C. A significant main effect was indicated for subtest. Using pairwise comparisons, results revealed that scores were statistically lower for the Filtered Words and Auditory Figure-Ground subtests than for the Competing Words and Competing Sentences subtests. The investigators also analyzed the Latino American participants' responses to test items using the Owens (1991) list of common Latino American phonological variations. The Competing Words and Filtered Words subtest needed the most dialectical scoring and all 20 Latino American participants required dialectical scoring.

The authors concluded, based on the composite score, that Latino American participants were 10% more likely to be placed into the disordered range when dialectical scoring was not applied; however, they more closely matched their Anglo American peers when dialectical scoring was considered. Bilingual participants also fell into the disordered range more often than the Anglo American participants when dialect was not considered, and were categorized as borderline when dialect was considered. The overall conclusion of the researchers is that dialectical scoring should be considered in children who are not Anglo American.

There were a few potential problems with this study. First, the age group of the participants was very limited and only included a small fraction of the ages of the children for whom this test was developed. Second, a comfortable loudness level was

used for presentation. If the presentation level of the signal was inadequately low in intensity, then the child may have had a more difficult time hearing the target words.

SCAN-A: A Test of Auditory Processing Disorders in Adolescents and Adults

A screening test for adolescents and adults was also developed by Keith (1995), because a comparable test was not available for this population, and APD is known to exist in this age group (Keith, 1994). Several versions of the SCAN-A were developed prior to the final version. After publishing the SCAN-A article (Keith, 1995), discrepancies between the information in the published article and the final SCAN-A test items were discovered for the Filtered Words and Auditory Figure-Ground subtests (Keith, 1998). The following review of the SCAN-A cites the filtering and signal-to-noise ratio values in the finalized version of the SCAN-A (Keith, 1998).

The SCAN-A: A Test of Auditory Processing Disorders in Adolescents and Adults (SCAN-A) is comprised of the same four subtests of the SCAN-C: Filtered Words, Auditory Figure-Ground, Competing Words, and Competing Sentences. The test was developed to be used for individuals 12 to 50 years old. Overall, the SCAN-A is comprised of more challenging listening conditions than the SCAN-C, to be more sensitive to central auditory processing problems in the older population for which it was designed.

To that end, there are two differences between the SCAN-A and SCAN-C. First, the Filtered Words subtest consists of words low-pass filtered at 1000 Hz in the SCAN-C and 750 Hz in the SCAN-A. A lower cutoff for the low-pass filter reduces the availability of some mid-frequency acoustic cues of speech, such as the second

formant. Second, the signal-to-babble ratio is +8 dB for the Auditory Figure-Ground test on the SCAN-C and +4 dB on the SCAN-A. A low signal-to-noise ratio indicates that the SCAN-A is presented with a higher level of noise compared to the SCAN-C. The instructions on the SCAN-A and SCAN-C vary somewhat to be appropriate for the target age group.

The test items in the SCAN-A are the same as those used for the SCAN-C (Keith, 2000). Each of the subtests is scored in the same way as the subtests of the SCAN and SCAN-C. For the Filtered Word and Auditory Figure-Ground subtests, the target word must be repeated verbatim to receive the test item correct.

During standardization of the SCAN-A, the test administrators consisted of 50 audiologists located at 21 different sites. One-hundred twenty-five participants (64 females and 61 males), ages 12 to 50 years, participated in the study. The study included 25 participants in each of the following age ranges: (a) 12-14; (b) 15-18; (c) 19-30; (d) 31-40; and (e) 41-50. The participants were 96% Caucasian, 12% African American, 8% Hispanic, and 3.2% from other ethnic backgrounds. Participants were required to have normal hearing (20 dB HL or better at 500, 1000, 2000, and 4000 Hz) and be native speakers of American English. Individuals with speech and language impairments or mental retardation could not participate. There was no mention of APD status. Testing of the SCAN-A was performed in a quiet room. Results in a sound-attenuating booth were not obtained, even though this test is used by audiologists in a sound attenuating-booth quite frequently. The presentation level of the SCAN-A was set at the participant's most comfortable level (MCL). The specific method of setting the MCL was not specified. The use of MCL as the

presentation level for a speech recognition test is known to produce highly variable results (Beattie & Warren, 1982; Stephens, Blegvad, & Krogh, 1977) and is not sufficient for a listener to achieve their maximum score (Beattie & Warren, 1982; Posner & Ventry, 1977). The participants were administered the SCAN-A in the subtest order described above. No randomization was performed.

After the standardization study was complete, standard scores for each of the SCAN-A subtests and the composite score were computed. No significant differences across all ages (12-50 years of age) arose for any of the subtest or composite scores of the SCAN-A. As a result of this finding, one standard score was developed for all age ranges of the SCAN-A.

Test-retest reliability data were analyzed on 38 participants, who were also part of the standardization study. Three age groups participated in this study. The age groups were 19-30 years of age, 31-40 years of age, and 41-50 years of age. There was no reference to the number of participants assigned to each age group. The Keith (1995) article does not report the test-retest reliability for participants aged 12 to 18 years. The second test administration occurred between one day and five months after the first administration, with a mean of 46 days. The test-retest reliability coefficient was 0.69 for the Total Test standard score, which is a moderately strong correlation. No analyses were performed for the subtests because of variability between test-retest scores (Keith, 1995). There is no mention if the test examiner was the same or different for both administrations of the SCAN-A.

Summary

APDs are one of the most misunderstood disorders in the field of audiology. Attempts have been made to produce a valid and reliable measure for diagnosing APDs in the school-aged population. However, no “gold standard” diagnostic test exists at the present time. Rather, a test battery approach is recommended for assessing APD (American Speech-Language-Hearing Association, 1996; Ferre & Wilber, 1986; Jerger & Musiek, 2000; Musiek & Chermak, 1994; Musiek et al., 1982; Singer et al., 1998).

Developing screening measures to identify children with a possible APD before an extensive test battery is performed also is important. However, at the present time, very few screening measures exist. The SCAN (Keith, 1986), SCAN-A (Keith, 1995) and the SCAN-C (Keith, 2000) were developed to be utilized as screening measures for children (SCAN and SCAN-C), and for adolescents and adults (SCAN-A). However, several studies report problems with the SCAN, including poor test-retest reliability, significant effects of test environment on the outcome scores, and learning effects (Amos & Humes, 1998; Emerson et al., 1997; Smith et al., 2000). Keith (2000) developed the SCAN-C to address some of these issues noted by other investigators. The SCAN-A (Keith, 1995) was developed before the SCAN-C and has not been revised. The test-retest reliability with participants 12-to-18-years of age was not addressed in the development of the SCAN-A. The effect of test environment also has not been investigated. The test was standardized in a quiet room but often is administered clinically in a sound-attenuating booth. The proposed study will examine the test-retest reliability and effect of test environment (sound-

attenuating booth versus quiet room) of the SCAN-A for the 12-to-15-year old age group. Test-retest reliability data and effect of test environment for this age group have yet not been reported, although this age group is possibly the largest target population for the SCAN-A. The SCAN-A is used frequently in attempting to screen adolescents for possible APDs. Often, these disorders are identified for the first time in adolescents. A high test-retest reliability is important in order to determine if an individual's results vary over time (Musiek & Chermak, 1994) and to interpret results accurately (Amos & Humes, 1998).

Research Questions

Three research questions are addressed in this study.

(1) Is there an effect of test administration time (day one versus day two) on adolescent performance on the SCAN-A?

(a) Is this effect significant for the composite score?

(b) Is this effect significant for each subtest score?

Rationale and Hypothesis for Question 1

To date, no studies have reported the test-retest reliability of the SCAN-A with adolescents ages 12-to-18 years of age. It is predicted that test-retest reliability will be low on the SCAN-A, reflecting possible learning effects. This prediction is based on the findings of previous investigators with the SCAN (Amos & Humes, 1998; Smith et al., 2000) This variability is expected because of the small number of test items for each subtest of the SCAN-A. The fewer number of test items within a measure, the more likely the outcome scores will be variable (Thorton & Raffin, 1978).

(2) Is there an effect of test environment (sound-attenuating booth versus quiet room) on the SCAN-A?

(a) Is this effect significant for the composite score?

(b) Is this effect significant for each subtest score?

Rationale and Hypothesis for Question 2

Performance is expected to be poorer in the quiet room than the sound-attenuating booth because ambient noise in the quiet room can cause auditory

distractions, as well as an inability to hear the test stimuli clearly. Testing is often conducted in a quiet school classroom; however, the school environment is a very noisy place with background noise levels in unoccupied classrooms ranging from 34.4 to 65.9 dBA, and a reverberation time of 0.2 to 1.27 sec (Knecht, Nelson, Whitelaw & Feth, 2002). Reverberation is defined as reflectance of sound waves as they come in contact with hard surfaces (Kurtovic, 1975), and reverberation time is defined specifically as the time it takes a sound to reduce 60 dB after the sound has ceased (Crandell & Smaldino, 2000). Reverberation can delay portions of the speech signal from being delivered to the intended individual directly, resulting in a smearing of the speech waveform and poor speech intelligibility. The typical noise levels and reverberation times of classrooms exceed the recommended levels required for optimal speech recognition. The recommended background noise levels range from 30-35 dBA and the recommended reverberation time ranges from 0.2 to 0.6 seconds (ASHA, 1995; Jerger & Musiek, 2000; Nelson, Soli, & Seltz, 2002). Testing students in a classroom will potentially make a speech understanding task quite difficult. One reason for the development of the SCAN-A was to have a tool that could be administered easily in a school environment (Keith, 1995). Therefore, the SCAN-A was standardized in a quiet room setting. Nevertheless, the test also is administered frequently in a sound-attenuating booth. Because of the potentially large difference in the ambient noise level of the quiet room compared to that of the sound-attenuating booth, it is possible that an individual would perform significantly better in a sound-attenuating booth because there will be fewer auditory distractions. If performance varies significantly depending on the test environment, then the test would have poor

reliability if it is administered in multiple locations. Hence, the effect of test environment consistency with the SCAN-A for the 12-to-15-year-old age group is critical to determine if this test can be utilized in both testing environments.

The results obtained in this proposed study are expected to demonstrate that the sound-attenuating booth environment produces higher subtest scores and composite scores than the quiet room environment. When listening in the sound-attenuating booth environment, the participant will be less distracted by outside noise, such as an air conditioner or individuals walking by the door.

(3) Are there interactions between these effects?

Rationale and Hypothesis for Question 3

It can be expected that in a quiet room that is not sound-treated, the noise levels will vary from day-to-day. As a result, reliability is expected to be poorer in the quiet room than in the sound-attenuating booth. This prediction suggests that there should be an interaction between the two main effects assessed in this study.

Method

Participants

The participants were 30 adolescents (19 girls and 11 boys) between the ages of 12-to-15 years (mean age of 13.27 years). The participants were recruited from the local community and the University of Maryland in the following ways: (a) word of mouth; (b) flyers at the University of Maryland; (c) advertisement on the Hearing and Speech Sciences website; (d) advertisements at local libraries and schools; and (e) advertisements posted in local stores (e.g., Giant). Volunteers were selected to participate if they were within this age range and exhibited normal hearing (20 dB HL or better from 250-8000 Hz), normal middle ear function (Shanks, Lilly, Margolis, Wiley, & Wilson, 1988; see explicit criteria in “Preliminary Testing” section), normal development for their age (using the Health, Education, and Behavioral Questionnaire), and were native speakers of standard American English. They were not invited to participate if they received special education services at school, had a known speech and/or language disorder, or had a known or suspected APD. The Health, Education, and Behavioral Questionnaire was used to determine if an individual met the criteria for inclusion in the study. Based on the questionnaire, an individual was not invited to participate because: (a) the child had a speech and/or language disorder at the current time of the investigation (question four); (b) the child’s dominant/primary language was not standard American English (question eight); and (c) the child had difficulty in one or more of the situations in question six.

These populations were not included because of the need to provide baseline, normative test-retest data for the SCAN-A.

Stimuli

A compact disc of the SCAN-A was obtained from Harcourt Assessment (www.harcourtassessment.com). The SCAN-A includes only one word list for each of the four subtests (Filtered Words, Auditory Figure-Ground, Competing Words, Competing Sentences). Therefore, to create multiple randomizations of each word list, the following process was completed. The calibration tone, instructions, practice words and target words for the four subtests of the SCAN-A were copied onto a laboratory computer and edited using Cool Edit software (Syntrillium Corporation). Each target word and the calibration tone were copied to individual files, and the maximum and average root mean square (RMS) values were analyzed. The average RMS value for all target words in a subtest was computed, as well as for the calibration tone provided with the test. This calibration tone was equivalent to the average maximum RMS level of the words. A total of 15 randomizations for each ear-specific list for all subtests was created (e.g., 15 randomizations were created for the Filtered Words right-ear tasks and 15 randomizations were created for the Filtered Words left-ear tasks). Each randomized list contains all of the target words for that ear-specific list for a subtest. After the randomizations were created for each subtest, eight Compact Discs (CD) (two for each subtest: one with the right-ear directed items and one with the left-ear directed items) were developed using CD burning software installed on a desktop personal computer. Each randomization was copied to a separate track on the CD for that subtest. Therefore, each CD contained 17 different

tracks, 15 tracks each containing one randomized list of stimuli, one track containing the instructions and practice items, and one track containing the calibration tone.

Procedure

Each participant was evaluated on several preliminary measures to determine if they satisfied the selection criteria. If a participant met the criteria, then experimental testing was performed. All testing was performed in the Department of Hearing and Speech Sciences in LeFrak Hall at the University of Maryland.

Preliminary testing. Participants were seated comfortably in a sound-attenuating booth and wore standard clinical E.A.R-3A 10 Ohm insert headphones. Each participant wore disposable ear tips that were discarded after the session. Preliminary audiometric testing in a sound-attenuating booth included standard routine audiometric measures, which consisted of air conduction threshold testing (250-8000 Hz), bone conduction testing (500-4000 Hz), and the presentation of Northwestern University Auditory Test Number Six (NU-6) (Tillman & Carhart, 1966) at a conversational level of 50 dB HL. A screen of middle ear function using the standard measure of tympanometry and ipsilateral and contralateral acoustic reflexes was performed. Tympanometry was considered within normal limits when the static admittance (Y) was between 0.3 and 1.70, the tympanometric peak pressure was greater than negative 100, and the equivalent ear canal volume was less than or equal to 2.0 cm³ (Margolis & Goycoolea, 1993; Wiley, Cruikshanks, Nondahl, Tweed, Klein, & Klein; Shanks et al., 1988). Ipsilateral and contralateral reflexes were measured at 500 Hz, 1000 Hz, and 2000 Hz and were considered within normal limits if the acoustic reflex threshold was below 100 dB HL (Handler & Margolis,

1977; Shanks et al., 1988). A questionnaire (Appendix A), which included questions about the participant's health, behavior, and education, was completed by each participant's guardian. The questionnaire helped to determine the participant's eligibility to participate in the research study. Participants who met the criteria were asked to participate in the experimental measures.

Prior to participating, the listeners and their guardians/parents were fully informed of the procedures and were given an opportunity to ask questions. Parents signed a Consent Form (Appendix B) and the adolescents signed an Assent Form (Appendix C). This investigation has been approved by the University of Maryland Institutional Review Board for Human Subjects Research.

Experimental procedures. The experimental procedures took place on two separate days, with the days approximately one month apart. A one month test-retest interval was used to keep the test-retest interval constant and to minimize the influence of maturation. A wide and non-constant test-retest interval was used during the SCAN-A standardization study (Keith, 1995). That test-retest interval ranged from 1 day to 5 months. In the current investigation, the participants listened through E.A.R. 3A 10 Ohm insert headphones to the test items of the SCAN-A CDs at a fixed presentation level of 60 dB SPL. This presentation level was confirmed before each experimental procedure using the calibration tone provided on each CD and a sound level meter. The test items were played through a Sony Compact Disk Radio Cassette Recorder Boom box (Model CFD-F10). On each experimental test day, the participant listened to all four subtests in two test environments: (a) a sound-attenuating booth at the University of Maryland Audiology Clinic, and (b) and a quiet

room (0135 LeFrak Hall) that had an ambient noise level of 50 dBA, similar to a school classroom. The target ambient noise level chosen for this experiment was taken from Knecht, Nelson, Whitelaw, and Feth (2002) which found that unoccupied classroom background noise levels ranged from 32 to 67 dBA in eight public school buildings in Ohio; however, there is no mention if these levels were measured when school was in session. To determine the appropriate level for this study, the average of 32 and 67 was taken, which resulted in a 50 dBA target level. The ambient noise level of the quiet room was measured before each test session and ranged from 47.2 to 50.3 dB SPL.

The order of presentation for the subtests of the SCAN-A was counterbalanced across participants. Additionally, the order of target words within each subtest was randomized. The room order (sound attenuating booth vs. quiet room) was counterbalanced across listeners.

All test items were presented at a comfortable listening level of 60 dB SPL. Prior to testing, the participant was asked to listen to the instructions provided on the SCAN-A Compact Disk. The participant then was asked if they understood what they were required to do. The standard practice words were then presented to the participant. Following the instructions and practice, the target words were presented to the participant. The participant repeated the word or sentence perceived, and the researcher, who was seated in the room with the participant, scored the response on the standard score sheets provided with the SCAN-A Compact Disk package. The participant's response was tape recorded for a subsequent internal reliability check. All scoring of the SCAN-A followed procedures described in the SCAN-A test

manual. These procedures will be described below. The participants did not receive feedback regarding the accuracy of their responses. The entire procedure took approximately two-and-a-half hours, with one-and-a-half hours the first day and one hour the second day.

Scoring Procedures. Participants' responses were scored using the score sheets provided with the SCAN-A. A test item was considered correct only if the entire item was repeated correctly. The total number of test items correct on a subtest is the raw score for that test. The maximum possible raw score for each subtest is 40 for the Filtered Words, 40 for the Auditory Figure-Ground, 60 for the Competing Words, and 20 for the Competing Sentences.

Standard scores were calculated for each subtest using the raw score and normative tables provided in the SCAN-A manual (Keith, 1994, p. 54). The standard scores are based on the transformation of linear z-scores (refer to Appendix D for more information). The maximum standard score for each subtest is 15 for the Filtered Words, 15 for the Auditory Figure-Ground, 15 for the Competing Words, and 12 for the Competing Sentences.

The total test standard score is calculated by summing the four subtest standard scores and then obtaining the relevant composite standard score using the norm table provided in the SCAN-A manual (Keith, 1994, p. 54). The composite standard score is based on a linear z-score transformation (Appendix D). The maximum possible composite score is 130.

Confidence intervals are provided for all standard scores. The confidence interval indicates a 68% level of confidence. For the Filtered Words, Auditory Figure-

Ground, and Competing Words the confidence interval is plus-or-minus two from the standard score. The confidence interval for the Competing Sentences is plus-or-minus one from the standard score.

Results

The means and standard deviations for the four subtests and the Total Test standard score of the SCAN-A for day one and day two booth and room scores are displayed in Table 1. The skewness and kurtosis for the four subtests and the Total Test standard score of the SCAN-A are displayed in Table 2. If most of the scores in a distribution are located at the lower end of the scale the distribution is defined as positively skewed. A negatively skewed distribution is a distribution where most of the scores are located at the higher end of the scale (Hinkle, Wiersma, & Jurs, 2003). The distributions in this study are all negatively skewed, indicating that the mean for each subtest and Total Test standard score was high.

The kurtosis of a distribution is defined as the degree of the peak of the distribution. A negative kurtosis value indicates a platykurtic distribution, meaning that the participant's scores are evenly distributed throughout the distribution. A leptokurtic distribution indicates that most of the participant's scores are located at the center of the distribution. A distribution is leptokurtic if the kurtosis value is positive. A distribution with zero kurtosis is considered mesokurtic (Hinkle, Wiersma, & Jurs, 2003). The distributions in this study are both leptokurtic and platykurtic.

Figure 1 displays the mean scores and standard deviations of all 30 participants for each condition (Booth Day One, Room Day One, Booth Day Two, Room Day Two) for the Filtered Words subtest. The highest possible standard score is 15 for the Filtered Words subtest. A two-way within-subjects repeated measures

Table 1

Means and Standard Deviations of Scores measured on Day 1 and Day 2 for the Four SCAN-A Subtests and Total Test Standard Score

Condition	SCAN-A Subtest				
	FW	AFG	CW	CS	TTSS
Booth, Day 1	8.03 (3.11)	7.87 (3.84)	8.2 (4.25)	9.7 (3.05)	88.1 (14.92)
Room, Day 1	8.53 (2.14)	7.6 (3.39)	8.3 (3.58)	9.26 (2.65)	88.6 (11.3)
Booth, Day 2	9.63 (1.93)	9.27 (3.11)	10.13 (3.31)	9.83 (2.35)	78.93 (11.06)
Room, Day 2	9.53 (2.26)	9.33 (3.29)	9.57 (3.24)	9.43 (2.81)	96.73 (14.84)

Note. Standard deviation is in parentheses; FW = Filtered Words subtest; AFG = Auditory Figure Ground subtest; CW = Competing Words subtest; CS = Competing Sentences subtest; TTSS = Total Test standard score.

Table 2

Skewness and Kurtosis Measured on Day 1 and Day 2 for the Four SCAN-A Subtests and the Total Test Standard Score

Condition	SCAN-A Subtest				
	FW	AFG	CW	CS	TTSS
Booth, Day 1	-.83 (.22)	-.44(-.97)	-.13(-1.07)	-1.23 (.99)	-.87 (1.14)
Room, Day 1	-.89 (.68)	-.46 (-.54)	-.55 (-.23)	-1.21 (2.10)	-.93 (.92)
Booth, Day 2	-.29 (-.78)	-.38 (-.27)	-.230 (-.96)	-1.8 (5.81)	-.336 (-.021)
Room, Day 2	-.9 (1.39)	-.11 (-1.49)	-.341 (-.55)	-1.04 (1.04)	-1.14 (1.05)

Note. Kurtosis is in parentheses; FW = Filtered Words subtest; AFG = Auditory Figure-Ground subtest; CW = Competing Words subtest; CS = Competing Sentences subtest; TTSS = Total Test standard score.

Filtered Words Subtest

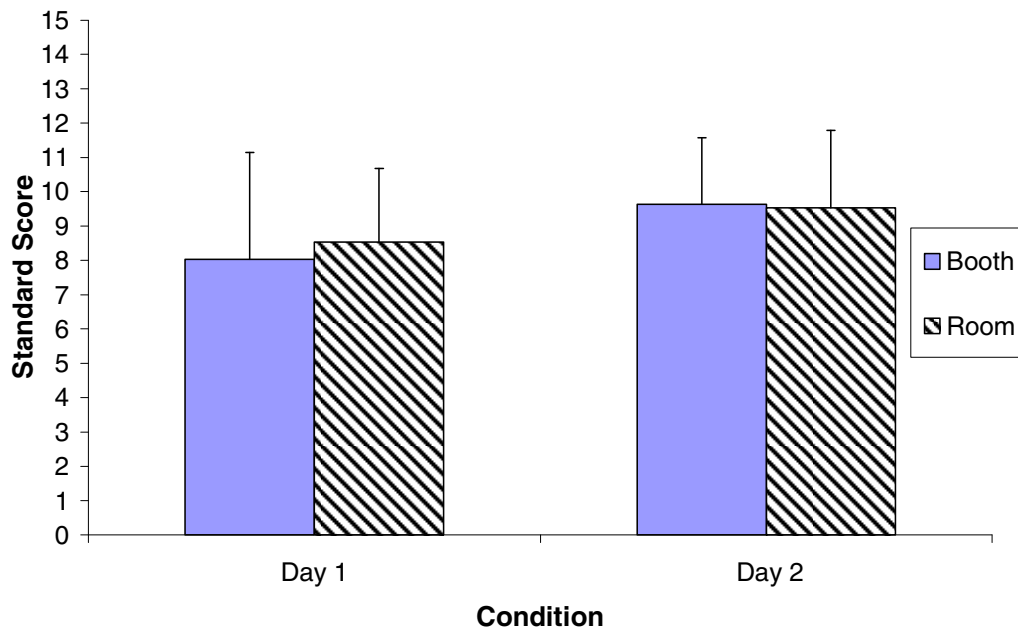


Figure 1. Means and standard deviations of the 30 participants for the Filtered Words subtest in each test room condition for each of the two test days.

analysis of variance (ANOVA) was conducted to evaluate the effect of test environment and test administration on the Filtered Words subtest. The within-subjects factors were test environment with two levels (booth and room) and test administration with two levels (day one and day two). There was a significant main effect of test administration [$F(1, 29) = 18.898, p < 0.01$], but no significant main effect of test environment [$F(1, 29) = 0.559, p > 0.05$]. There was no significant interaction between test administration and test environment [$F(1, 29) = 0.923, p > 0.05$]. Thus, scores on day two were significantly higher than scores on day one for the Filtered Words subtest standard scores.

The Auditory Figure-Ground data were analyzed using ANOVA with a repeated measures design with two within-subjects variables (test administration and test environment). The means and standard deviations of the 30 participants are shown in Figure 2. The highest standard score that can be achieved on this subtest is 15. There was a significant main effect of test administration [$F(1, 29) = 12.608, p < 0.01$], but no significant main effect of test environment [$F(1, 29) = 0.063, p > 0.05$] and no significant interaction between test environment and test administration [$F(1, 29) = 0.246, p > 0.05$]. A comparison of the mean scores from the two test administration days indicates that the Auditory Figure-Ground subtest standard scores are significantly higher on day two when compared to day one.

Means and standard deviations for the Competing Words subtest standard scores are displayed in Figure 3. The highest standard score possible is 15. Results for the Competing Words subtest standard scores, using a within-subjects repeated measures ANOVA, revealed a significant main effect of test administration

Auditory Figure-Ground Subtest

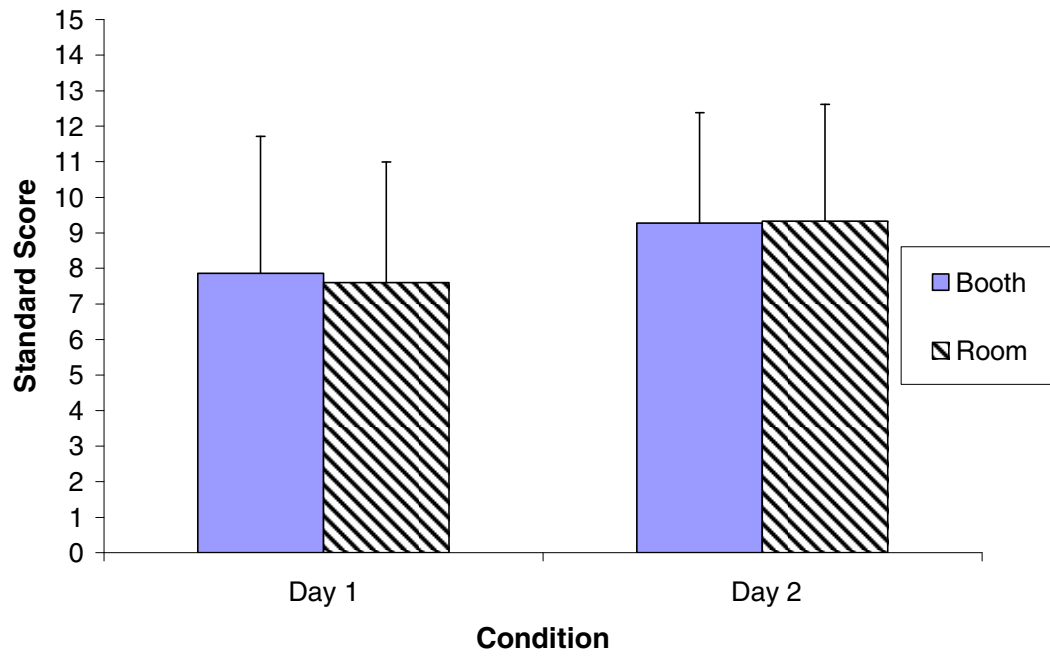


Figure 2. Means and standard deviations of the 30 participants for the Auditory Figure-Ground subtest for each room test condition on each of the two test days.

Competing Words Subtest

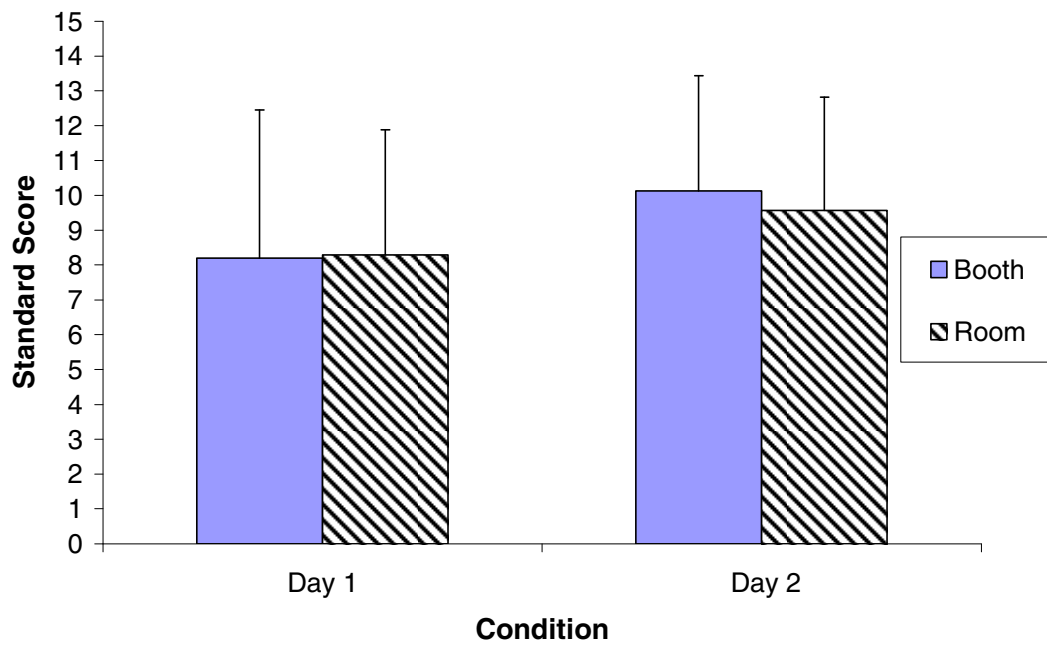


Figure 3. Means and standard deviations of the 30 participants for the Competing Words subtest for each room condition on each of the two test days.

[$F(1, 29) = 46.207, p < 0.01$]. There was not a significant main effect of test environment [$F(1, 29) = 0.589, p > 0.05$] nor a significant interaction of test administration and test environment [$F(1, 29) = 0.890, p > 0.05$]. The repeated measures ANOVA results reveal that the Auditory Figure-Ground subtest standard scores are significantly higher on day two than on day one.

The final subtest of the SCAN-A is Competing Sentences. Figure 4 displays the means and standard deviations for the Competing Sentences subtest standard scores. The highest possible standard score for the Competing Sentences subtest is 12. A two-way within-subjects repeated measures ANOVA was conducted to determine the effect of test administration and test environment on the Competing Sentences subtest standard scores. There were no significant main effects of test administration [$F(1, 29) = 0.145, p > 0.05$], or of test environment [$F(1, 29) = 2.454, p > 0.05$], and no significant interaction of test administration and test environment [$F(1, 29) = 0.004, p > 0.05$]. Total test standard scores were calculated for each participant by summing the scores from the four subtests.

Figure 5 displays the means and standard deviations for the Total Test standard score. The highest Total Test standard score that can be achieved is 130. A two-way within-subjects repeated measures ANOVA revealed a significant main effect of test administration [$F(1, 29) = 41.925, p < 0.01$]. There was no significant main effect of test environment [$F(1, 29) = 0.572, p > 0.05$] and no significant interaction between test administration and test environment [$F(1, 29) = 0.632, p > 0.05$]. The Total Test standard scores were significantly higher on day two than on day one.

Competing Sentences Subtest

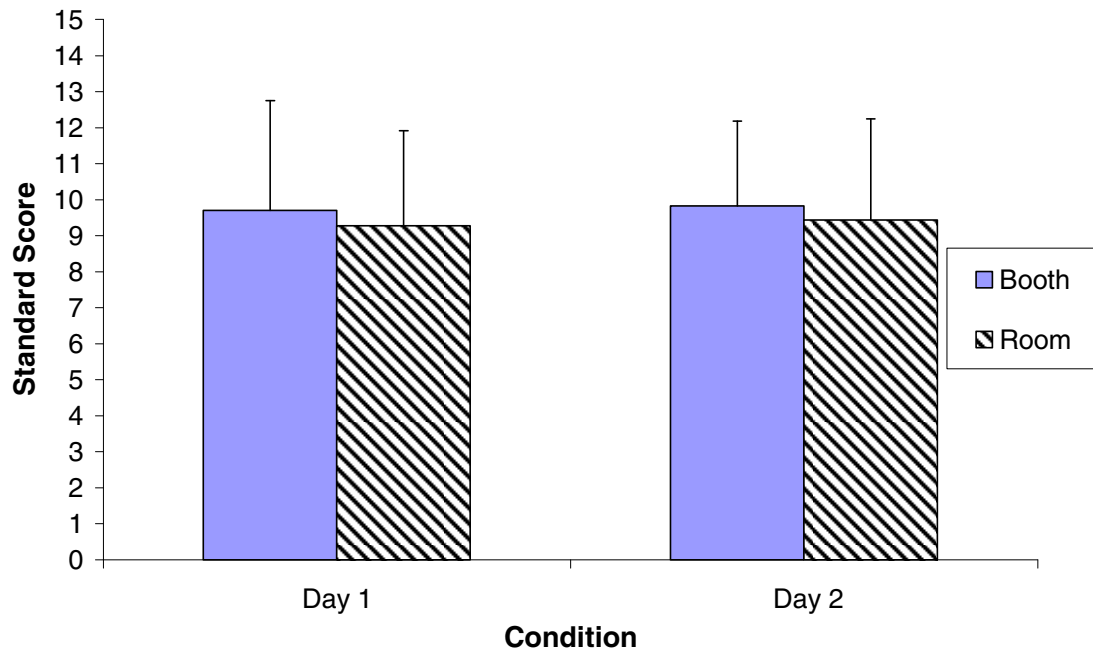


Figure 4. Means and standard deviations for the Competing Sentences subtest for the 30 participants for each room condition each of the two test days.

Total Test Standard Score

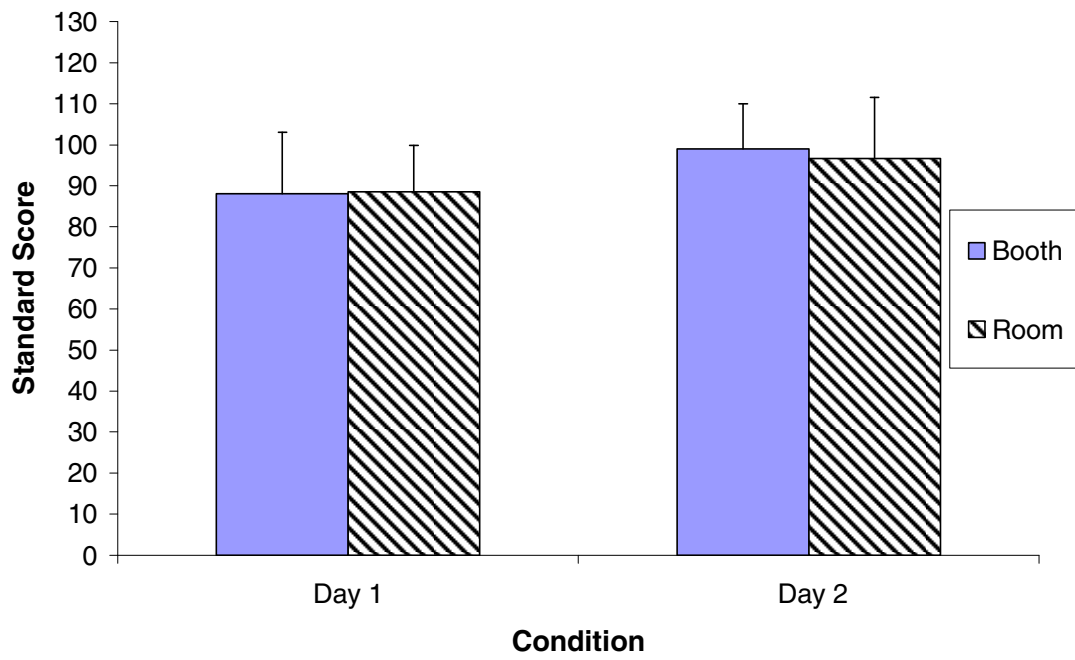


Figure 5. Means and standard deviations for the Total Test standard score for each test room condition for each of the two test days.

Pearson Product-moment correlations were conducted between test administrations to facilitate comparison to data reported by other investigators. The correlations are displayed in Table 3 and indicate a low-to-moderate correlation for the Competing Sentences subtest, a moderate correlation for the Filtered Words, Auditory Figure-Ground, and Total Test standard scores, and a moderate-to-high correlation for the Competing Words subtest (Hinkle, Wiersma, & Jurs, 2003). A subsequent analysis was performed to evaluate the change in performance over the four test administrations, regardless of test room. This analysis was possible because results of the two-way ANOVA for the Total Test standard score revealed no significant main effect of test environment. Therefore, it was possible to analyze the data by test administration order regardless of the test environment. The means and standard deviations of the Total Test standard score by test administration are presented in Figure 6. A one-way ANOVA was conducted for the within-subjects factor of administration order (administration one, administration two, administration three, and administration four). The dependent variable was the Total Test standard score. The results for the ANOVA indicate a significant effect of test administration order [$F(1, 29) = 18.592, p < 0.01$].

Post-hoc testing, using the Bonferroni approach, was conducted to evaluate pairwise differences among the means. There was a significant difference between the means for administration one and administration three ($p < 0.05$) and a significant difference between the means for administration one and administration four ($p < 0.05$). No significant differences were found between administration one and

Table 3

Pearson Product-moment Test-retest Correlations of Scores measured on Day 1 and Day2 for the Four SCAN-A Subtests and the Total Test Standard Score

Condition	SCAN-A Subtest				
	FW	AFG	CW	CS	TTSS
Room	.509**	.553**	.741**	.627**	.621**
Booth	.579**	.668**	.865**	.497**	.554**

Note. FW = Filtered Words subtest; AFG = Auditory Figure-Ground subtest; CW = Competing Words subtest; CS = Competing Sentences subtest; TTSS = Total Test standard score.

** $p < .01$

Total Test Standard Scores by Administration

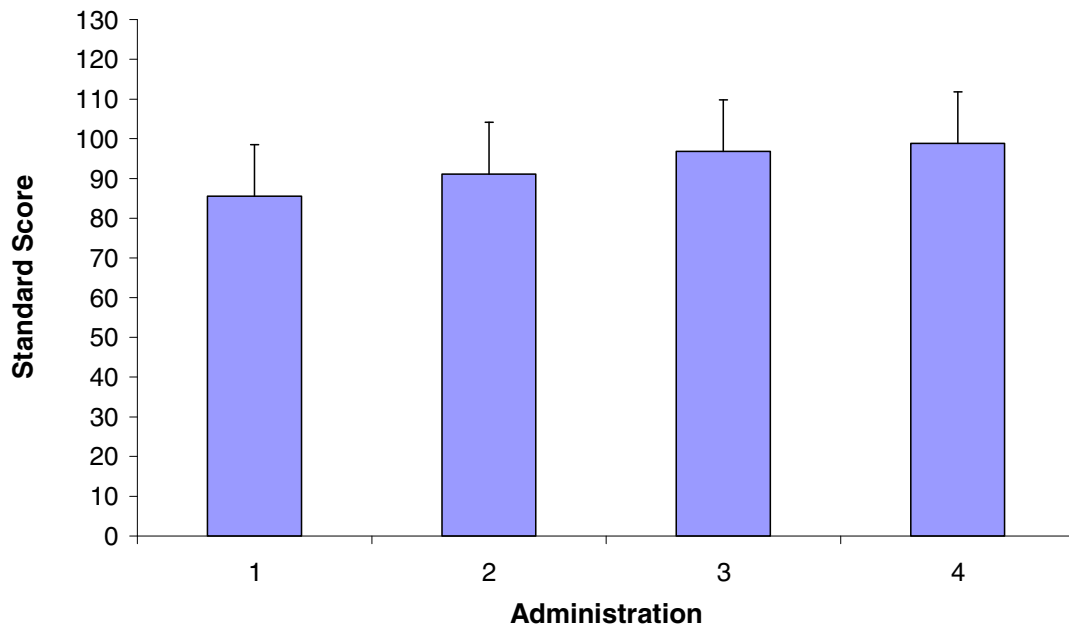


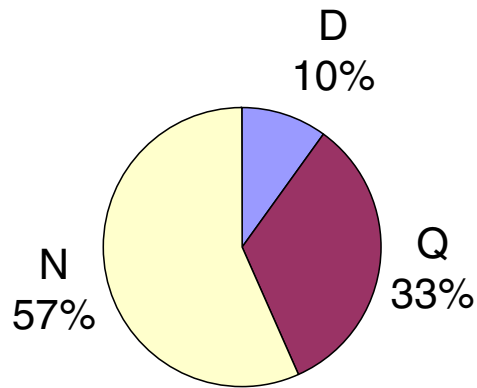
Figure 6. Means and standard deviations of the Total Test standard score by test administration order (regardless of test environment).

administration two, between administration two and administration three, between administration two and administration four, and between administration three and administration four. The first test administration score was significantly lower (worse) when compared to both test administrations three and four. It should be noted that administrations one and two were on the same day, as were administrations three and four. These two test days were one month apart.

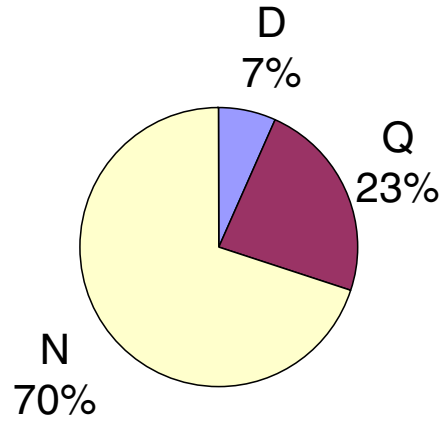
Based on the Total Test standard score, participants were assigned to one of three SCAN-A categories: within normal limits, questionable results, and disordered. These categories are intended to help to determine if individuals should be referred for further APD testing, and are the three possible outcomes of the SCAN-A screening procedure. Figures 7a, 7b, 7c, and 7d display the number and percentage of individuals assigned to each category for the four test administrations. A participant was designated as disordered if their Total Test standard score was less than or equal to 70, following recommendations in the SCAN-A test manual. A participant was designated as questionable if their Total Test standard score ranged from 71-to-85. Participants were assigned into the normal category if their Total Test standard score was greater than 85.

The specificity, or the ability for a test to identify correctly participants who do not have APD, should be 100% for all test administrations considering that participants in this study were normally developing with no concerns of APD. For this study, sensitivity was the percentage of participants (out of 30 participants) who were placed in the normal category. The false positive rate is the percentage of participants (out of 30 participants) who were placed in the disordered or questionable

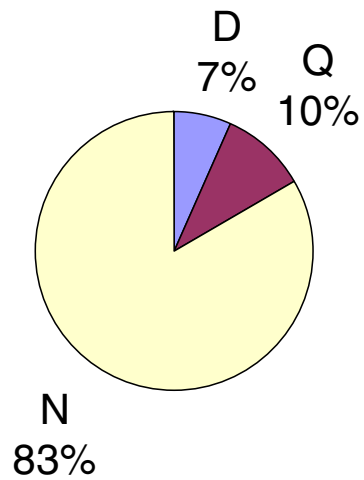
a.) **Administration 1**



b.) **Administration 2**



c.) **Administration 3**



d.) **Administration 4**

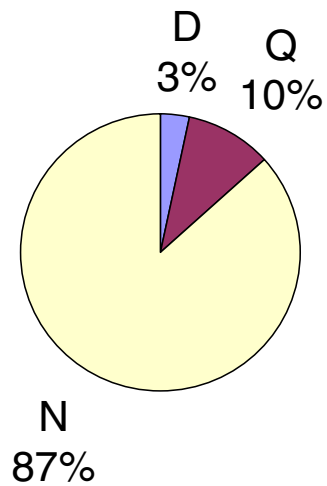


Figure 7. Percentage of participants categorized as disordered (D), questionable (Q), or normal (N) using the Total Test standard score for test each test administration one, two, three, and four (regardless of test environment).

categories. Calculations of specificity rates revealed that there is low specificity (57%) and a high false positive rate (43%) for administration one. Administration two has a higher specificity (70%) and a low false positive rate (30%) but these results are inadequate. For administrations three and four, the specificity is higher (83% and 87%, respectively) and the false positive rate is lower (17% and 13%, respectively) when compared to administration one and administration two. Administration one categorized the most individuals as either disordered or questionable.

Three participants were categorized as disordered and 10 individuals were categorized as questionable with the first test administration. In this first administration, nearly half of the children did not exhibit clearly normal results. With increasing test administrations, fewer individuals were classified as disordered and questionable, with only one individual designated as disordered and three individuals designated as questionable by administration four.

Discussion

Effect of Test Administration

A significant test-retest effect was found when comparing the scores obtained on day one to day two for the Filtered Words subtest, Auditory Figure-Ground subtest, Competing Words subtest, and Total Test standard score measured separately in the quiet room and sound-attenuating booth. These results are similar to the results with younger children on the SCAN reported by Amos and Humes (1998) and Smith et al. (2000). Amos and Humes (1998) investigated the test-retest reliability of the SCAN with first- and third-graders. Results revealed that participants exhibited a significant learning effect for the Filtered Words, Competing Words, and Total Test standard score. Smith et al. (2000) investigated test-retest reliability with 5-to-10-year olds on the SCAN. The children showed a learning effect for all subtests and the Total Test standard score. Although these previous studies were evaluating test-retest reliability with the original SCAN, the SCAN-A used in the current study is not substantially different from the SCAN, because many of the same test items are included in both versions. However, the current investigation is the first to evaluate test-retest reliability and effect of test environment of the SCAN-A with adolescents ages 12-to-15 years. This study also is the only investigation, other than the standardization study (Keith, 1995), to investigate the SCAN-A.

The only subtest that did not display a significant effect of test-retest was the Competing Sentences subtest. This subtest has the fewest number of test items

($n=20$) and most participants performed maximally (18 to 20 items correct) with this subtest on administration one, indicating a ceiling effect.

A measure has adequate test-retest reliability when the instrument has a strong correlation coefficient of $r > 0.8$ and no significant differences between the means (Amos & Humes, 1998; Nunnally, 1959). Based on the correlation analysis and ANOVA results, all subtests and the Total Test standard score do not meet both criteria. Therefore, based on the results of this investigation, the test-retest reliability of the SCAN-A does not appear to be adequate for the 12-to-15 year old age group.

The test-retest results reported in the current study are inconsistent with the results reported by Keith (1995). Keith (1995) found no significant difference between test-retest scores for the Total Test standard score with the SCAN-A; however, the Total Test standard score correlation was $r = 0.69$, which is less than the recommended correlation of $r > 0.8$ (Amos & Humes, 1998). The correlations for the Total Test standard score for this current study were $r = .621$ and $r = .554$ for the room and booth, respectively, and are lower than the reported correlation for the standardization study (Keith, 1995). The differences between the results reported across studies could be associated with the difference in age groups that were investigated, the test-retest interval used, or the analysis reported.

The age group in the current study is younger (12-to-15 years of age) than the age group (19 to 50 years of age) in the Keith (1995) study. The sample size in the standardization study consisted of three age groups (19-30, 31-40, and 41-50 years of age) with a total of 38 participants; however, there is no mention of the number of participants per age group nor the average age in each group (Keith, 1995).

Moreover, separate data for each age group were not reported. It is likely that differences in results between the present study and that of Keith (1995) are associated with the differences in participants' ages in the two studies. Individuals who are older may be able to attend more easily to the stimuli for a longer period of time, have more experience with test taking, and be more familiar with the test items. Children's performance with a listening task has been shown to be more variable than that of adults (Allen, Wightman, Kisler, & Dolan, 1989). Children also have been shown to have shorter durations of attention (Allen & Wightman, 1994; Allen & Wightman, 1995) when compared to adults. This variability in performance and shorter duration of attention is seen primarily for young children (pre-school to elementary school-age), and these abilities become adult-like in the adolescent years (Kidd & Hogben, 2006). It is possible that the adolescents in the current study were still developing these cognitive skills that appear to influence performance on challenging listening tasks, and hence, have not yet reached the level of performance reported by Keith (1995).

The test-retest interval used in the standardization study was not controlled, and ranged from 1 day to 5 months, with an average of 46 days (Keith, 1995). In the current study, the test-retest interval for all participants was constant at approximately one month.

Because there was extreme inconsistency between test-retest subtest scores of the SCAN-A in the standardization study, correlations were only conducted for the Total Test standard score (Keith, 1995). The current study's test-retest differences were analyzed for all four subtests as well as the Total Test standard score. If an

analysis of test-retest was performed for each subtest in the standardization study (Keith, 1995), results may have revealed significant differences between test-retest scores for some of the subtests. Overall, the findings from this current study show that when performance is not optimal on a given subtest, there is a significant effect of test-retest performance on the SCAN-A results for the 12-to-15 year old age group.

This significant effect of test-retest performance on the SCAN-A suggest a learning effect between day one and day two for the Filtered Words, Auditory Figure-Ground, Competing Words and the Total Test standard score. Participants, on average, performed better on day one versus day two for three of the four subtests and the Total Test standard score.

Effect of Test Environment

No effect of test environment was found for any of the subtests or Total Test standard score. This result does not support the preliminary study by Emerson et al. (1997), which observed a possible difference in test environments for the Auditory Figure-Ground and Total Test standard score for five of their six participants. The previous study observed that performance was poorer in the quiet room compared to the sound booth, although data were not analyzed statistically. The discrepancy across investigations could be a result of differences in age, randomization protocols, participant population, test-retest intervals used, stimulus presentation level, number of participants, room characteristics, and test items. The participants in the current study were adolescents, ages 12-to-15, whereas the participants in the Emerson et al. (1997) study were between the ages of 5.9 and 11.8 years of age. Individuals, less than 13 years of age, may have a harder time attending to speech stimuli in a noisy

and reverberant classroom environment (Nabelek & Nabelek, 1994). The ability of individuals to recognize consonants in a reverberant room becomes adult-like by approximately 13 years of age (Neuman & Hochberg, 1983).

The current study counterbalanced the subtests and randomized the test items, whereas Emerson et al. (1997) did not randomize or counterbalance these factors. Without randomization, the participants may have possibly performed better on administration two versus administration one because they were familiar with the test item order. Additionally, the test-retest interval for the current study was one month and the previous study's test-retest interval was one week (Emerson et al., 1997). With such a short test-retest interval, children may be able to remember the test items more easily, especially because the test items were not randomized in the previous study.

The participants who were recruited for the Emerson et al. (1997) investigation were clients of a speech-language pathologist. These children may have had a speech production problem, and therefore, the test administrator may have had a difficult time interpreting their answers to some of the test items. If the administrator had difficulty interpreting the participant's answers, the participant's recorded score and true score may be significantly different. There is concern that current APD test measures may stimulate and actually measure language areas of the brain (Poremba et al., 2003; Salvi et al., 2002). If the children in the Emerson et al. (1997) study had language deficits, their poor performance on the Auditory Figure-Ground subtest, as well as the Total Test standard score could possibly be a result of their speech and language disorder, and not an APD.

The presentation level for the Emerson et al. (1997) study was not reported. It is likely that the volume was adjusted individually to a comfortable listening level, following instructions in the SCAN manual. The test items may not have been presented at the same level for the booth and the room conditions. Without one set level, or a description of how the presentation level was determined, the differences in scores between environments could be a result of differing presentation levels. The volume would most likely need to be set at a higher level in the quiet room than the sound-attenuating booth.

The Emerson et al. (1997) study was a preliminary study with only six participants. The current study had a larger sample size (30 participants). It is unclear whether the test site would have had a significant effect if a larger number of participants were tested in the Emerson et al. (1997) study.

Finally, the previous study utilized the SCAN, whereas the current study used the SCAN-A. There are differences in filtering and speech-to-noise ratios, for the Filtered Words and Competing Words subtests, respectively when comparing the SCAN to the SCAN-A.

Total Test Standard Score by Administration

Because there was no effect of test environment, results for the Total Test standard score were collapsed across test locations to examine the effect of test administration over time in more detail. The one-way ANOVA for test administration revealed that participants performed better (had a higher score) on administrations three and four compared to administration one.

There are several possible explanations for these results. First, practice effects could have occurred because participants, after one test administration, understood what they were required to do. This helped the participants perform better on subsequent administrations. After the first test administration, participants knew what to expect for each subtest, they understood the directions, and they knew the approximate test length. Second, a learning effect may have occurred between administration one and administration three or four because there is only one list of test items. After a few test administrations, participants heard the same test items multiple times and were able to determine more accurately the correct answer. During this investigation, participants were not provided with the correct answer. Presumably, this strategy reduced possible learning effects. Finally, a significant effect was found between administration one and administration three or four possibly because of maturational effects that occur during adolescence. Maturation differences have been found for individuals as old as 12 years-of-age when compared to adults for consonant recognition (Hazan & Barrett, 2000). Adolescents as old as 15-years-of-age have been found to perform more poorly in noisy or reverberant environments when compared to adults for auditory identification tasks (Johnson, 2000). Late auditory evoked potentials, which assess the auditory cortex (Scherg, Vasjar, & Picton, 1989; Vaughn & Ritter, 1970), show developmental changes throughout adolescence (Courchesne, 1978; Ponton, Don, Eggermont, Waring, & Masuda, 1996). Thus, although the test-retest interval was short (1 month) in the current investigation, it is possible that auditory maturation in the central auditory pathways affected the performance of the study participants.

Based on these results, it is difficult to determine which test administration is the most accurate in determining a participant's auditory processing ability. It can be hypothesized that the first test administration is not accurate in determining auditory processing ability, because a significant proportion of participants, who were reportedly developing normally, were diagnosed as disordered or questionable. However, it is not ideal or time efficient to administer the SCAN-A more than once or twice to determine if a full APD test battery is warranted.

Future research should be directed at determining the principal source of performance improvement with repeated administrations of the SCAN-A. One strategy is to determine if learning effects occur with the SCAN-A by testing participants twice with the same word list on the same day. If an individual's score significantly increases between test administration one and test administration two, the results would indicate the participant learned the test items. Practice effects could be identified by administering two different test item lists on the same day. If the participant's score on administration two is significantly higher, then a practice effect has occurred. To determine if maturational effects occur, the screening test could be administered at different test-retest intervals (e.g., one day, one week, one month, one year). If significant differences occur between test administration days, then maturation has occurred (this is true as long as the test has adequate test-retest ability). The findings of these studies will have important implications for improving development of screening measures for APD. For example, if participants show a learning effect for test items, then they could be trained on the words prior to testing. If participants show a significant practice effect, then they may require some practice

prior to test administration. Finally, if maturation is significant, then correction factors would be needed in order to compare the first test administration to the second test administration.

SCAN-A Outcomes

The SCAN-A categorizes individuals into one of three groups: disordered, questionable, and normal. These categories were developed to help the clinician to determine an appropriate course of action for each patient. If an individual falls into the disordered or questionable category, the results warrant further APD testing. Based on the results of the current study, 13 participants (three disordered and 10 questionable), or 43% of the participant total, would be placed into a category that would warrant a full APD test battery following the first test administration. However, with subsequent test administrations, fewer and fewer participants would require a full APD test battery. By administration three, only five participants (17% percent of the sample size) would be referred for a full APD test battery and following administration four, only four participants (13% of the participant total) would be referred for a full APD battery. These results are remarkable given that all participants in the current study were normally developing, with no parental or school concerns of an APD.

The specificity should be 100% for all test administrations because the participant population in this study were all normally developing with no concerns of APD; however, this is not the case for this study. Acceptable sensitivity and specificity rates for hearing loss screening measures are 90-95% (Roeser & Northern,

1981). Based on this acceptable specificity rate, none of the four test administrations, with specificity rates ranging from 57% to 87%, produced acceptable results.

An analysis of individual performance over the four test administrations indicated that most of the participants who were placed in the disordered or questionable category on administration one, were placed in the normal category by administration three or four. The significant difference between administration one and administration three and four suggests a learning effect. A learning effect would indicate that the participants remembered, or learned, the test items between administration one and administration three and four. Although this suggests a possible learning effect, this performance pattern may also reflect practice or maturation. Three participants were placed in the disordered or questionable category for all four test administrations. This indicates that practice, learning, and/or maturation effects did not occur for these three participants because their scores did not greatly improve between test administration one and test administration four.

The SCAN-A is a screening measure; therefore, based on the results of this measure, individuals should be referred for further evaluation if their scores fall into the disordered or questionable categories. However, anecdotally and from personal experience, the SCAN-A is used frequently as a diagnostic measure. Patients have been identified as having an APD with the administration of the SCAN-C or SCAN-A only. Also, an APD may be identified for the first time during adolescence. This is a troubling occurrence because individuals who do not display common characteristics of APD are categorized as having an APD based on the SCAN-A alone. This use of the SCAN-A likely results in a high cost to parents, to the schools,

and possibly affects family dynamics and expectations. Parents of adolescents who are identified by the SCAN-A as having a possible APD incur time and monetary costs for their adolescent to have a full auditory processing evaluation. A full auditory processing battery takes, on average, two to three hours, depending on which tests are administered. Parents must spend approximately half of their work day to have their adolescent evaluated for APD, as well as spend money on an APD test battery, which may not be covered fully by insurance. The parents and adolescent may feel anxiety waiting for the full APD test battery and may assume that the adolescent probably has an APD, when, in fact, a large number of individuals are misidentified with the SCAN-A.

Many children and adolescents are evaluated for an APD because they are not performing well in school. Individuals may not perform well in school for many reasons, which include attention deficit disorder, learning disability, poor vision, poor hearing, emotional problems, and/or an auditory processing disorder. When the SCAN-A is used to identify an individual with an APD, the adolescent subsequently may receive special services at school. These services include FM systems and preferential seating in the classroom, which represent a high monetary cost to the school system, and single out the adolescent. These special services will likely not be adequate for adolescents who do not have an APD. With a misdiagnosis of an APD, the individual may not be referred for testing of other possible causes of the individual's difficulty in school. The true disorder may be attention deficit disorder, learning disability, and/or poor vision, etc.

Limitations of the Current Study

Limitations of the current study exist. One limitation is that only one test-retest interval of one month was used. A short test-retest interval was desirable for participant retention. Nevertheless, this interval may not have been short enough. An alternative strategy is to test two groups of adolescents: one with a relatively brief test-retest interval (e.g., one week), and one with a relatively long test retest interval (e.g., one month). This protocol would permit an assessment of the separate effects of maturation versus practice and learning effects.

Another limitation is that the participant population consisted exclusively of individuals who were normally developing. Although the current study demonstrated a test-retest effect, future research should investigate test-retest reliability with a population of individuals with APD. At the current time, this may be a challenging task because there is no gold standard to identify an individual with an APD.

Although a classroom with a noise level of approximately 50 dBA was utilized for the current investigation (Knecht et al, 2002), the classroom used may not have been sufficiently representative of a room in a primary or secondary school. Most of the results of the current investigation were obtained in the summer and on the weekends, and therefore, may not be representative of noise and distractions that would be present during the school day.

Finally, this investigation did not address the content and construct validity of the SCAN-A. In this case, content validity is the ability of the test items to assess an individual's overall auditory processing ability. Construct validity is defined as the ability of the test to measure what it is supposed to measure. However, because the

SCAN-A is not reliable from one test administration to the next with 12-to-15-year olds, the concepts of content and construct validity may be of less importance than modifying the test to improve reliability at least for this age group.

Conclusions

The results of the current study indicate that the SCAN-A has low specificity, a high false positive rate, and poor test-retest reliability with adolescents 12-to-15 years of age. Future research should attempt to unravel the principal source of the performance improvement observed between day one and day two in the current investigation, and capitalize on the findings to develop a more reliable screening measure with high specificity, a low false positive rate and acceptable test-retest reliability for the adolescent age group. An APD screening measure with the desired requirements is important because a full APD test battery is time consuming (about two hours in length) to administer and over-referrals carry a high cost for parents and educational institutions. Further research also should investigate the test-retest reliability of the SCAN-A with the 16-to-18-year old age group and with a group of individuals diagnosed with APDs.

Appendix A

Health, Education, and Behavior Questionnaire
Parent/Guardian

Name of Participant: _____
Address: _____
Participant Number: _____ Telephone: _____
Age: _____ Email: _____
Person Completing this Form: _____
Date of Birth: _____

1. Has your child ever received any of the following special services at school?
- | | |
|--|---|
| <input type="checkbox"/> Reading Specialist | <input type="checkbox"/> Occupational Therapy |
| <input type="checkbox"/> English to Speakers of Other Languages (ESOL) | <input type="checkbox"/> Physical Therapy |
| <input type="checkbox"/> Speech and Language Services | <input type="checkbox"/> Instructional Aide |
| <input type="checkbox"/> Other | <input type="checkbox"/> One-on-one Aide |

If other, please explain:

2. Has your child repeated a grade? **Yes No**
If yes, please explain:

3. In what grade is your child enrolled? _____

4. Does your child have any speech and/or language disorders at the present time (e.g., stuttering, articulation)? **Yes No**
If yes, please explain:

5. Has your child ever had any surgery on their ears? **Yes No**
If yes, please explain:

6. Does your child have a history of chronic/recurrent ear infections? **Yes No**
If yes, please indicate the approximate number of ear infections and age:

7. Does your child have difficulty with any of the following situations? **Yes No**

If yes, please circle which situation and explain below:

- a.) Difficulty in hearing and/or understanding speech in the presence of background noise
- b.) Difficulty in understanding speech which is degraded (e.g., muffled speech, fast speech)
- c.) Difficulty in following instructions which are spoken
- d.) Difficulty in discriminating and identifying speech sounds
- e.) Inconsistent responses to auditory information

8. What languages does your child speak at home and at school? _____
9. What is the child's dominant/primary language? _____

Appendix B

Parental Permission Form

Project Title	<i>Reliability of Testing Environment of the Scan-A with Children Ages 12-15</i>
Why is this research being done?	<i>This research project is being conducted by Dr. Sandra Gordon-Salant and Michele Spencer in the Department of Hearing and Speech Sciences at the University of Maryland, College Park. We are inviting your child to participate in this research project because he or she is between 12 and 15 years of age, and has normal hearing. The purposes of the research project are (1) to determine if scores are consistent from one test administration to the next on the SCAN-A, a clinical test that assesses the ability of individuals to comprehend words and sentences in difficult listening situations; and (2) to determine if different test environments (sound proof versus quiet room) change the score on the SCAN-A.</i>
What will I be asked to do?	<i>The procedures involve two one-hour sessions, one month apart, where your child will be seated in a hearing test booth and/or quiet room and asked to listen to words and sentences under headphones. The words and sentences may be muffled, two words or sentences may be presented at the same time, or the words may be presented with noise in the background. Your child will be required to give a verbal response to each word or sentence heard. Prior to these procedures you will be asked to complete a Health Checklist about your child. Your child will also undergo a routine hearing test that will last for approximately one hour. Your child can receive student service learning credit (3 hours) and/or knowledge regarding his/her hearing ability.</i>
What about confidentiality?	<i>We will do our best to keep your personal information confidential. To help protect you and your child's confidentiality, all data will be kept in a locked file cabinet in 0119 LeFrak Hall in the Department of Hearing and Speech Sciences. If we write a report or article about this research project, your identity, as well as your child's identity will be protected to the maximum extent possible. Your information will be shared with representatives of the University of Maryland, College Park or governmental authorities if you or someone else is in danger or if we are required to do so by law.</i>
What are the risks of this research?	<i>There are no known risks associated with participating in this research.</i>
What are the benefits of this research?	<i>This research is not designed to help your child personally, but the results may help the investigator learn more about the consistency of scores on the SCAN-A and the effect of test environment differences (sound attenuating booth versus quiet room) on the SCAN-A. We hope that, in the future, other people might benefit from this study through improved understanding of the SCAN-A.</i>
Do I have to be in this research? May I stop participating at any time?	<i>Your child's participation in this research is completely voluntary. You may choose for your child not to take part at all. If you decide to permit your child to participate in this research you may stop him/her from participating at any time. If you decline to permit your child to participate in this study or if you stop their participation at any time, you will not be penalized or lose any benefits to which you otherwise qualify.</i>

Project Title	<i>Reliability of Testing Environment of the SCAN-A with Children Ages 12-15</i>	
What if I have questions?	<p><i>This research is being conducted by Dr. Sandra Gordon-Salant in the Department of Hearing and Speech Sciences at the University of Maryland, College Park. If you have any questions about the research study itself please contact Dr. Sandra Gordon-Salant at:</i></p> <p><i>0100 LeFrak Hall, College Park, MD 20742</i> <i>(301) 405-4225</i> <i>sgordon@hesp.umd.edu</i></p> <p><i>If you have any questions about your rights as a research subject or wish to report a research-related injury, please contact: Institutional Review Board Office, University of Maryland, College Park, MD 20742; (e-mail) irb@deans.umd.edu; (telephone) 301-405-0678</i></p> <p><i>This research has been reviewed according to the University of Maryland, College Park IRB procedures for research involving human subjects.</i></p>	
Statement of Age of Subject and Consent	<p><i>Your signature indicates that:</i></p> <p><i>you are at least 18 years of age;</i></p> <p><i>your child is 12 to 17 years of age;</i></p> <p><i>the research has been explained to you;</i></p> <p><i>your questions have been fully answered;</i></p> <p><i>you freely and voluntarily choose for your child to participate in this research project.</i></p>	
Signature and Date	Name of Child	
	Name of Parent	
	Signature of Parent	
	Date	

Appendix C

Assent Form

Project Title	<i>Are the Scores on the SCAN-A the Same on Different Days and in Different Rooms?</i>
Why is this research being done?	<i>The goal of this study is to find out if scores from a listening test are the same from one time to the next time, and if taking the test in two different rooms changes the score. We want you to be in the study because you are 12 to 15 years old and have normal hearing. The test is often used to assess hearing problems in students who have difficulty hearing in the classroom.</i>
What will I be asked to do?	<i>The test involves two days, one month apart, where you will sit in two different rooms and will be asked to listen to words and sentences with headphones. The words and sentences may sound muffled, the words and sentences may be heard at the same time, or the words may be hard to hear because of noise. You will tell the researcher what you hear after each word and sentence is presented. Before these tests, you will also have your hearing checked. The whole experiment will take about 3 hours. For helping with the study you can earn student service learning credit (3 hours) and/or knowledge regarding your hearing ability.</i>
What about confidentiality?	<i>Information about you will be kept secret and your name will not used.</i>
What are the risks of this research?	<i>There are no known risks associated with participating in this research.</i>
What are the benefits of this research?	<i>This study will help the researchers to determine if this listening test is a good test to use with teenagers. The only benefit to you is to find out more about your hearing ability.</i>
Do I have to be in this research? May I stop participating at any time?	<i>You understand that you are free to ask questions and you can withdraw from the study at any time without penalty.</i>
Contact Information of Investigators	<i>Sandra Gordon-Salant, Ph.D. Department of Hearing and Speech Sciences University of Maryland, College Park (301) 405-4225 sgordon@hesp.umd.edu Michele Spencer Department of Hearing and Speech Sciences University of Maryland, College Park (301) 405-7454 m Spencer@hesp.umd.edu</i>
Statement of Consent	<i>You indicate that: you agree to be in this study</i>
	<i>Name of Child</i>
	<i>Date</i>

Appendix D

Derivation of Standard Scores for the SCAN-A

The following is taken from the SCAN-A manual:

Subtest standard scores used in the SCAN-A were derived from linear z-scores, because it is unreasonable to assume that the traits being measured by the SCAN-A are normally distributed in the population. Linear z-scores were transformed to a standard score scale with a mean of 10 and a standard deviation of three. The SCAN-A Total Test standard scores were derived from linear z-scores obtained from the distribution of the sum of the subtest standard scores. The linear z-scores for the Total Test scores were then transformed to a standard score scale with a mean of 100 and a standard deviation of 15. (Keith, 1994, p. 42)

References

- Allen, P., & Wightman, F. (1994). Psychometric functions for children's detection of tones in noise. *Journal of Speech and Hearing Research, 37*, 205-215.
- Allen, P., & Wightman, F. (1995). Effects of signal and masker uncertainty on children's detection. *Journal of Speech and Hearing Research, 38*, 503-511.
- Allen, P., Wightman, F., Kisler, D., & Dolan, T. (1989). Frequency resolution in children. *Journal of Speech and Hearing Research, 32*, 317-322.
- Altepeter, T. S. (1989). The PPVT-R as a measure of psycholinguistic functioning: a caution. *Journal of Clinical Psychology, 45*, 935-941.
- American Speech-Language-Hearing Association (1995). Acoustics in educational settings. *American Speech-Language-Hearing Association, 37*, 15-19.
- American Speech-Language-Hearing Association (1996). Central auditory processing: current status of research and implications for clinical practice. *American Journal of Audiology, 5*, 41-54.
- Amos, N. E., & Humes, L. E. (1998). SCAN test-retest reliability for first- and third-grade children. *Journal of Speech, Language, and Hearing Research, 41*, 834-845.
- Baker, H., & Leland, B. (1968). *Detroit Tests of Learning Aptitude*. Indianapolis: Bobbs-Merrill.
- Beasley, D. S., & Maki, J. (1976). Time and frequency altered speech. In N. Lass (Ed.). *Contemporary issues in experimental phonetics* (pp. 419-458). New York: Academic Press.

- Beattie, R. C., & Warren, V. G. (1982). Relationships among speech threshold, loudness discomfort, comfortable loudness, and PB max in the elderly hearing-impaired. *The American Journal of Otology*, 3, 353-358.
- Bellis, T. J., (1997). *Understanding auditory processing disorders in children*. Retrieved July 22, 2007, from [http://www.asha.org/public/hearing disorders/understand-apd-child.htm](http://www.asha.org/public/hearing_disorders/understand-apd-child.htm)
- Bellis, T. J., & Burke, J. R. (1996). Overview of central tests. In T. J. Bellis (Ed.). *assessment and management of central auditory processing disorders in the educational setting* (113-139). San Diego: Singular Publishing Group.
- Bench, J., & Bamford, J. (1979). *Speech-hearing tests and the spoken language of hearing-impaired children*. London: Academic Press.
- Berlin, C. I., Cullen, J. K., Hughes, L. F., Berlin, J. L., & Lowe-Bell, S. S. (1975). Dichotic processing of speech. In M. D. Sullivan (Ed.). *Central auditory processing disorders: proceedings of a conference at the university of nebraska medical center, Omaha* (pp. 36-46).
- Bocca, E. (1958). Clinical aspects of cortical deafness. *Laryngoscope*, 68, 301-309.
- Bocca, E., Calero, C., & Cassinari, V. (1954). A new method for testing hearing in temporal lobe tumours. *Acta Otolaryngological*, 44, 219-221.
- Bocca, E., Calero, C., Cassinari, V., & Migliavacca, F. (1955). Testing cortical hearing in temporal lobe tumours. *Acta Otolaryngological*, 42, 289-304.
- Cacace, A. T., & McFarland, D. J. (1995). Opening pandora's box: The reliability of CAPD tests. *American Journal of Audiology*, 4, 61-62.

- Cacace, A. T., & McFarland, D. J. (1998). Central auditory processing disorders in school-aged children: A critical review. *Journal of Speech, Language, and Hearing Research, 41*, 355-373.
- Calearo, C., & Antonelli, A. R. (1968). Audiometric findings in brainstem lesions. *Acta Otolaryngologica, 66*, 305-319.
- Chermak, G.D. (1996). Central Testing. In S.E. Gerber (Ed.), *The handbook of pediatric audiology* (pp. 206-253). Washington, DC: Gallaudet University Press.
- Chermak, G. D., & Musiek, F. E. (1992). Managing central auditory processing disorder in children and youth. *American Journal of Audiology, 1*, 61-65.
- Chermak, G. D., & Musiek, F. E. (1997). Central auditory processing disorders: New perspectives. San Diego: Singular Publishing Group.
- Cherry, R. (1980). *Selective Auditory Attention Test-Manual*. St. Louis: Auditec.
- Clark, J. G. (1980). Central auditory dysfunction in school children: A complication of management suggestions. *Language, speech, and hearing services in the schools, 11*, 208-213.
- Cullen, J. K., & Thompson, C. L. (1974). Masking release for speech in subjects with temporal lob resections. *Archives of Otolaryngology, 100*, 113-116.
- Courchesne, E. (1979). Neurophysiological correlates of cognitive development: Changes in long-latency event-related potentials from childhood to adulthood. *Electroencephalography and Clinical Neurophysiology, 45*, 468-482.

- Cordes, A. K. (1994). The reliability of observational data: I. theories and methods for speech-language pathology. *Journal of Speech and Hearing Research*, 37, 264-278.
- Crandell, C. C., & Smaldino, J. J. (2000). Room acoustics for listeners with normal-hearing and hearing impairment. In M. Valente, H. Hosford-Dunn, & Roeser, R. J. (Eds.), *Audiology Treatment* (pp. 601-638). New York: Thieme.
- DiSomoni, F. (1978). *The Token Test for Children*. Boston: Teach Resource Corporation.
- Domitz, D. M., & Schow, R. L. (2000). A new CAPD battery- Multiple auditory processing assessment: Factor analysis and comparisons with SCAN. *American Journal of Audiology*, 9, 1-11.
- Dunn, L. M., & Dunn, L. M. (1981). *Peabody Picture Vocabulary Test-Revised*. Circle Pines, Minnesota: American Guidance Services.
- Elliott, L., & Katz, D. (1979). *Development of Northwestern University Children's Perception of Speech (NU-CHIPS) test*. St Louis: Auditec.
- Emerson, M. F., Crandall, K. K., Seikel, J. A., & Chermak, G. D. (1997). Observations on the use of the SCAN to identify children at risk for central auditory processing disorder. *Language, Speech, and Hearing Services in Schools*, 28, 43-49.
- Ferre, J. M., & Wilber, L. A. (1986). Normal and learning disabled children's central auditory processing skills: An experimental test battery. *Ear and Hearing*, 7, 336-343.

- Goldman, R., Fristoe, M., & Woodcock, R. (1976). *Goldman-Fristoe-Woodcock Auditory Skills Test Battery*. Circle Pines, Minnesota: American Guidance Services.
- Goldstein, R. (1961). Hearing and speech follow-up after left hemispherectomy. *Journal of Speech and Hearing Disorders, 26*, 126-129.
- Goldstein, R. Goodman, A. C., & King, R. B. (1956). Hearing and speech in infantile hemiplegia before and after left hemispherectomy. *Neurology, 6*, 869-875.
- Goodglass, H. (1967). Binaural digit presentation and early lateral brain damage. *Cortex, 3*, 295-306.
- Handler, S. D., & Margolis, R. H. (1977). Predicting hearing loss from stapedial reflex thresholds in patients with sensorineural impairment. *Transactions of the American Academy of Ophthalmology and Otology, 84*, 425-431.
- Haskins, H. (1949). *A phonetically balanced test of speech discrimination for children*. Unpublished master's thesis, Northwestern University, Evanston, Illinois.
- Hazan, V., & Barrett, S. (2000). The development of phonemic categorization in children aged 6-12. *Journal of Phonetics, 28*, 377-396.
- Hinkle, D. E., Wiersma, W., & Jurs, S. G. (2003). *Applied statistics for the behavioral sciences*. Boston: Houghton Mifflin Company.
- Hodgson, W. (1967). Audiological report of a patient with left hemispherectomy. *Journal of Speech and Hearing Disorders, 32*, 39-45.
- Jerger, J. (1960). Observations on auditory behavior in lesions of the central auditory pathways. *Archives of Otolaryngology, 71*, 65-74.

- Jerger, S. (1983). Decision matrix and information theory analysis in the evaluation of neuroaudiologic tests. *Seminars in Hearing, 4*, 121-132.
- Jerger, J., & Musiek, F. (2000). Report of the consensus conference on the diagnosis of auditory processing disorders in school-aged children. *Journal of the American Academy of Audiology, 11*, 467-474.
- Jerger, J. F., & Jerger, S. W. (1974). Auditory findings in brainstem disorders. *Archives of Otolaryngology, 99*, 342-349.
- Jerger, J. F., & Jerger, S. W. (1975). Clinical validity of central auditory tests. *Scandinavian Audiology, 4*, 147-163.
- Johnson, C. E. (2000). Children's phoneme identification in reverberation and noise. *Journal of Speech, Language, and Hearing Research, 43*, 144-157.
- Keith, R. W. (1977). *Central auditory dysfunction*. New York: Grune & Stratton.
- Keith, R. W. (1986). *SCAN: A Screening Test for Auditory Processing Disorders*. San Antonio, TX: The Psychological Corporation.
- Keith, R. W. (1994). *SCAN-A: A Test for Auditory Processing Disorders in Adolescents and Adults*. San Antonio, TX: The Psychological Corporation.
- Keith, R. W. (1995). Development and standardization of SCAN-A: test of auditory processing disorders in adolescents and adults. *Journal of the Academy of Audiology, 6*, 286-292.
- Keith, R. W. (1998). Development of SCAN-A: Test of auditory processing disorders in adolescents and adults [Letter to the Editor], *Journal of the American Academy of Audiology*, p. 311.

- Keith, R. W. (2000). Development and standardization of SCAN-C test for auditory processing disorders in children. *Journal of the American Academy of Audiology, 11*, 438-445.
- Kidd, J. C., & Hogben, J. H. (2006). Development of auditory saltation and its relationship to reading and phonological processing. *Journal of Speech, Language, and Hearing Research, 49*, 352-366.
- Kimura, D. (1961). Some effects of temporal lobe damage on auditory perception. *Canadian Journal of Psychology, 15*, 166-171.
- Kirk, S., McCarthy, J., & Kirk, W. (1968). *Illinois Test of Psycholinguistic Abilities*. Champaign: University of Illinois Press.
- Knecht, H., Nelson, P., Whitelaw, G., & Feth, L. (2002). Structural variables and their relationship to background noise levels and reverberation times in unoccupied classrooms. *American Journal of Audiology, 11*, 2.
- Korsan-Bengsten, M. (1973). Distorted speech audiometry: A methodological and clinical study. *Acta Otolaryngological, (suppl. 310)*, 1-75.
- Kurtovic, H. (1975). The influence of reflected sound upon speech intelligibility. *Acoustica, 33*, 32-39.
- Lynn, G. E., & Gilroy, J. (1972). Neuro-audiological abnormalities in patients with temporal lobe tumors. *Journal of Neuroscience, 17*, 167-184.
- Lynn, G. E., & Gilroy, J. (1977). Evaluation of central auditory dysfunction in patients with neurological disorders. In R. W. Keith (Ed.). *Central auditory dysfunction* (pp. 177-221). New York: Grune & Stratton.

- Margolis, R. H., & Goycoolea, H. G. (1993). Multifrequency tympanometry in normal adults. *Ear and Hearing, 14*, 408-413.
- Marriage, J., King, J., Briggs, J., & Lutman, M. E. (2001). The reliability of the SCAN test: Results from a primary school population in the UK. *British Journal of Audiology, 35*, 199-208.
- McFarland, D. J., & Cacace, A. T. (1995). Modality specificity as a criterion for diagnosing central auditory processing disorders. *American Journal of Audiology, 4*, 36-48.
- Moray, N. (1959). Attention in dichotic listening: Affective cues and the influence of instructions. *Quarterly Journal of Experimental Psychology, 11*, 56-60.
- Musiek, F. E. (1983a). Results of three dichotic speech tests with intracranial lesions. *Ear and Hearing, 8*, 317-323.
- Musiek, F. E. (1983b). Assessment of central auditory dysfunction: The dichotic digits test revisited. *Ear and Hearing, 4*, 79-83.
- Musiek, F. E. (1999). Central auditory tests. *Scandinavian Audiology, 28*(Suppl. 58), 33-46.
- Musiek, F. E., Baran, J. A., & Pinheiro, M. L. (1990). Duration of pattern recognition in normal subjects and patients with cerebral and cochlear lesions. *Audiology, 29*, 304-313.
- Musiek, F. E., Bellis, T. J., & Chermak, G. D. (2005). Nonmodularity of the central auditory nervous system: implications for (central) auditory processing disorder. *American Journal of Audiology, 14*, 128-138.

- Musiek, F. E., & Chermak, G. D. (1994). Three commonly asked questions about central auditory processing disorders: Assessment. *American Journal of Audiology, 3*, 23-27.
- Musiek, F. E., Geurkink, N. A., & Kietel, S. A. (1982). Test battery assessment of auditory perceptual dysfunction in children. *Laryngoscope, 92*, 251-257.
- Musiek, F. E., Gollegly, K., Kibbe, K., & Verkest-Lenz, S. (1991). Proposed screening tests for central auditory disorders: Follow-up on the dichotic digits test. *American Journal of Otology, 12*, 109-113.
- Musiek, F. E., & Pinherio, M. (1985). Dichotic speech tests in the detection of central auditory dysfunction. In M. Pinherio & F. Musiek (Eds.), *Assessment of central auditory dysfunction: Foundations and clinical correlates* (pp. 201-218). Baltimore: Williams and Wilkins.
- Musiek, F. E., Wilson, D. H., Pinheiro, M. L. (1979). Audiological manifestations in split-brain patients. *Journal of the American Audiological Society, 5*, 25-29.
- Myklebust, H. (1954). *Auditory disorders in children*. New York: Grune & Stratton.
- Nabelek, A. K., & Nabelek, I. V. (1994). Room acoustics and speech perception. In J. Katz (Ed.), *Handbook of clinical audiology* (pp. 624-637). Baltimore: Williams and Wilkins Company.
- Nelson, P. B., Soli, S. D., & Seltz, A. (2002). *Classroom acoustics II: Acoustical barriers to learning*. Melville, New York: Acoustical Society of America.
- Neuman, A. C., & Hochberg, I. (1983). Children's perception of speech in reverberation. *Journal of the Acoustical Society of America, 71*, 2145-2149.

- Nunnally, J. C., Jr. (1959). *Tests and measurements: Assessment and prediction*. New York: McGraw-Hill.
- Olsen, W. O., Noffsinger, D., & Kurdziel, S. (1975). Speech discrimination in noise by patients with peripheral and central lesions. *Acta Otolaryngological*, 80, 375-382.
- Owens, R. E. (1991). *Language disorders: A functional approach to assessment and intervention*. New York: Macmillan.
- Oxbury, J., & Oxbury, S. (1969). Effect of temporal lobectomy on the report of dichotically presented digits. *Cortex*, 5, 3-14.
- Pinherio, M. L. (1977). Tests of central auditory function in children with learning disabilities. In R. Keith (Ed.), *Central auditory dysfunction* (pp.223-256). New York: Grune and Stratton.
- Pinherio, M. L., & Ptacek, P. H. (1971). Reversals in the perception of noise and tone patterns. *Journal of the Acoustical Society of America*, 49, 1778-1782.
- Ponton, C. W., Don, M., Eggermont, J. J., Waring, M. D., & Masuda, A. (1996). Maturation of human cortical auditory function: Differences between normal-hearing children and children with cochlear implants. *Ear and Hearing*, 17, 430-437.
- Poremba, A., Saunders, R. C., Crane, A. M., Cook, M., Sokoloff, L., & Mishkin, M. (2003). Functional mapping of the primate auditory system, *Science*, 299, 568-572.

- Posner, J., & Ventry, I. M. (1977). Relationships between comfortable loudness levels for speech and speech discrimination in sensorineural hearing loss. *The Journal of Speech and Hearing Disorders, 42*, 370-375.
- Ptacek, P. H., & Pinherio, M. L. (1971). Pattern reversal in auditory perception. *Journal of the Acoustical Society of America, 49*, 493-498.
- Rintelman, W. (1985). Monaural speech tests in the detection of central auditory disorders. In F. E. Musiek & M. L. Pinherio (Eds.), *Assessment of central auditory dysfunction: Foundation and clinical correlates* (pp. 173-200). Baltimore: Williams and Wilkins
- Roeser, R. R. & Northern, J. L. (1981). Screening for hearing loss and middle ear disorders. In R. R. Roeser & M. P. Downs (Eds.), *Auditory disorders in school children* (pp. 120-150). New York: Thieme.
- Rosenweig, M. (1951). Representation of the two ears at the auditory cortex. *American Journal of Physiology, 167*, 147-158.
- Salvi, R. J., Lockwood, A. H., Frisina, R. D., Coad, M. L., Wack, D. S., & Frisina, D. R. (2002). PET imaging of the normal human auditory system: Responses to speech in quiet and in background noise. *Hearing Research, 170*, 96-106.
- Scherg, M., Vasjar, J., & Picton, T. (1989). A source analysis of the late human auditory evoked potentials. *Journal of Cognitive Neuroscience, 1*, 336-355.
- Schultz M. C. (1972). *Analysis of clinical behavior in speech and hearing*. Englewood Cliffs, New Jersey: Prentice-Hall.
- Shanks, J. E., Lilly, D. J., Margolis, R. H., Wiley, T. L., & Wilson, R. H. (1988). Tympanometry: ASHA working group on aural acoustic-immittance

- measurements committee on audiologic evaluation. *Journal of Speech and Hearing Disorders*, 53, 354-377.
- Singer, J., Hurley, R. M., & Preece, J. P. (1998). Effectiveness of central auditory processing tests with children. *American Journal of Audiology*, 7, 73-84.
- Smith, T. T., Bradham, T., Chandler, L., & Wells, C. (2000). The effect of examiner's race on the performance of African American children on the SCAN. *Language, Speech, and Hearing Services in Schools*, 31, 116-125.
- Spache, G. D. (1953). A new readability formula for primary-grade reading materials. *Elementary School Journal*, 53, 410-413.
- Speaks, C. (1980). Central auditory tests. In M. M. Paparella & D. A. Schumrick (Eds.), *Otolaryngology, Volume II: The ear* (pp. 1846-1860). Philadelphia: W. B. Saunders.
- Stephens, S. D. G., Blegvad, B., & Krogh, H. J. (1977). The value of some suprathreshold auditory measures. *Scandinavian Audiology*, 6, 213-221.
- Thorton, A. R., & Raffin, M. J. M. (1978). Speech-discrimination scores modeled as a binomial variable. *Journal of Speech and Hearing Research*, 21, 507-518.
- Tillman, T. W., & Carhart R. (1966). *An Expanded Test for Speech Discrimination Utilizing CNC Monosyllabic words*. Northwestern University Auditory Test No. 6. Brooks Air Force Base, TX, USAF School of Aerospace Medical Technical Report.
- Treisman, A. M., & Geffen, G. (1968). Selective attention and cerebral dominance in perceiving and responding to speech messages. *Quarterly Journal of Experimental Psychology*, 79, 139-150.

- Vaughn, H. G., & Ritter, W. (1970). The sources of auditory evoked responses recorded from the human scalp. *Electroencephalography and Clinical Neurophysiology*, 28, 360-367.
- Weschler, D. (1974). *Weschler Intelligence Scale for Children-Revised*. New York: Psychological Corporation.
- Wiley, T. L., Cruikshanks, K. J., Nondahl, D. M., Tweed, T. S., Klein, R., & Klein, B. E. K. (1996). Tympanometric measures of older adults. *Journal of the American Academy of Audiology*, 7, 260-268.
- Willeford, J. A. (1977). Assessing central auditory behavior in children: A test battery approach. In R. Keith (Ed.), *Central auditory dysfunction* (pp. 43-72). New York: Grune and Stratton.
- Woods, A. G., Pena, E. D., & Martin, F. N. (2004). Exploring possible sociocultural bias on the SCAN-C. *American Journal of Audiology*, 13, 173-184.