Article

A New CAPD Battery—Multiple Auditory Processing Assessment: Factor Analysis and Comparisons with SCAN

Deborah M. Domitz
Ronald L. Schow
Idaho State University, Pocatello

Based on recommendations by Musiek & Chermak (1994, *American Journal of Audiology, 3,* 23–27) and ASHA (1996, *American Journal of Audiology, 5*(2), 41–54), a battery of four commonly used tests was selected and recorded for use in assessing school children. These tests were labeled the Multiple Auditory Processing Assessment (MAPA) and then administered to an initial sample of 81 third grade children, along with the SCAN screening test for auditory processing disorders. Afterward, several exploratory factor analyses were performed on the findings, and comparisons were made between the results for the MAPA and SCAN. Four separate factors emerged from the four MAPA tests, which were linked closely to the components of central auditory processing disorders (CAPDs) defined by ASHA (1996, *American Journal of Audiology, 5*(2), 41–54). These factors were labeled monaural separation/closure (MSC), auditory pattern/temporal ordering, binaural integration, and binaural separation (BS). SCAN appears to measure two of these factors, MSC and BS. Use of MAPA is encouraging. Our findings suggest it may provide an appropriate multiple-test, CAPD battery for third grade children, and it meets at least some of the objectives described in the consensus document proposed by ASHA (1996, *American Journal of Audiology, 5*(2), 41–54).

Key Words: central auditory processing disorders, screening, factor analysis, SCAN, sensitivity, specificity, CAPD, diagnosis, Multiple Auditory Processing Assessment (MAPA)

According to ASHA (1996), central auditory processing disorders (CAPDs) are observed deficiencies in one or more of the following six auditory behaviors: sound localization and lateralization, auditory discrimination, auditory pattern recognition, temporal aspects of audition (resolution, masking, integration, ordering), auditory performance decrements with competing acoustic signals, and auditory performance decrements with degraded acoustic signals. These behaviors are assumed to apply to verbal and nonverbal stimuli and are assessed by tests in five areas including temporal processes, localization and lateralization, low-redundancy monaural speech, dichotic (binaural) stimuli, and binaural interaction. Thus, screening tools and diagnostic tests must be able to assess a variety of problems associated with CAPD. We are entirely supportive of the general direction of this ASHA position paper, but a companion article suggests nomenclature that allows a more direct linkage between the six auditory behaviors specified by ASHA and the five areas of testing (Schow, Seikel, Chermak, & Berent, 2000). We emphasize that when the ASHA document notes that one or more of these problems constitute CAPD, it is clear that CAPD is not considered to be a syndrome in which a given set of problems is always present, but rather the presence of any one problem can constitute CAPD. This point is underscored by recent publications in which tests over several discrete areas are recommended for CAPD assessment (Chermak & Musiek, 1997; Musiek & Chermak, 1994; Musiek & Pinheiro, 1987; Schow & Chermak, 1999).

One major problem is the absence of a "gold standard" in CAPD assessment. It is rare that discrete neurological lesions allow definite CAPD diagnosis (Jerger, Johnson, & Loiselle, 1988) and in most children CAPD is suspected on the basis of behavioral and observational data but in the absence of discrete lesions. Chermak & Musiek (1997) suggest that neurodevelopmental disorder underlies CAPD in 65 to 70% of diagnosed CAPD, although it is beyond the capability of most clinics to confirm in the majority of cases. If, as inferred by ASHA (1996), multiple problem areas need to be assessed, then establishing a gold standard is even more complicated. Methods that identify certain processing problems may be insensitive to other areas of dysfunction, and not all persons judged to have CAPD will score poorly on all measures of dysfunction (Musiek & Lamb, 1994; Willeford, 1985). Notwithstanding, Singer, Hurley, & Preece (1998) have recently recommended the use of clinical decision analysis (CDA) used in site of lesion testing as a method to determine CAPD test efficacy for an individual test or test battery. On the basis of their findings they recommend the use of an abbreviated battery in the identification of CAPD in children, involving only one or two tests. Given the range of processing deficits that may constitute CAPD and the relative insensitivity of particular tests to sample all areas of dysfunction, it would seem that if the ASHA (1996) document is correct, it is unlikely that one test could constitute the diagnostic standard nor provide adequate direction for intervention. A simple version of CDA drawn from notions of a single site of lesion or a uniprocessing CAPD entity would not seem to provide much help in CAPD diagnosis if we are dealing with a multiprocessing disorder.

One more complication is evident in that many diagnostic tests were initially designed to identify the site-of-lesion in neurologically impaired adults, yet are currently used in assessing functional proficiency in children. Thus, professionals need to select an age- and language-level-appropriate series of tests for children (Musiek & Chermak, 1994; Willeford, 1985).

Despite these difficulties, the ASHA (1996) document states that two priorities for research in CAPD include establishing sensitivity and specificity of assessment tools as well as guidelines for identification of children at risk for CAPD. In response to this challenge, the current work is an effort to use the ASHA (1996) consensus document and some current common assessment tools to explore these issues. Our strategy was to examine the range of pertinent auditory abilities in a cross section of third grade children, using a battery of tests, with a special focus on statistically evaluating the underlying factor structure.

A screening test for auditory processing disorders (SCAN) (Keith, 1986) is commonly used as a multitest screening measure of CAPD. Keith (1995) notes that the purpose of SCAN is to determine possible central auditory nervous system (CANS) disorders, identify auditory processing problems, and identify children who may benefit from remediation. SCAN, therefore, is considered useful for both screening and diagnosis. SCAN has recently been issued in an updated version (Keith, 2000), but the original version was used in this study. Despite having three subtests, results from all three have recently been shown to constitute only two definite factors based on factor analysis (Schow & Chermak, 1999). Chermak, Styer, & Seikel (1995) compared the screening utility of SCAN and the Selective Auditory Attention Test (SAAT) (Cherry, 1980) and found that the two tools do not necessarily identify the same children as being at risk for CAPD, though overlap did occur. SAAT identified a larger number of children than SCAN as being at risk and had better agreement with parental reports. We conclude that further information is needed regarding the sensitivity and specificity of SCAN compared with a battery of tests that are tailored to the ASHA (1996) recommendations.

Musiek & Chermak (1994) recommended use of a battery of four tests to examine all levels of the CANS in children. This approach is reasonable for diagnosis of the diffuse neurodevelomental form of CAPD common in children. This battery approach was selected for more careful study, factor analysis, and comparison with SCAN because of the battery compatibility with the ASHA (1996) document and because of its suitability and recommendation for use with children.

Methods

Participants

All participants were selected from third grade classrooms in three local, public elementary schools because this is considered an appropriate age for screening CAPD and an age when children are often referred for CAPD testing. Participation was dependent on grade level and parental response to a consent form that explained the purpose of the study. Participants were tagged if their school records included information regarding a prior diagnosis of speech/language, learning, or attention disorders. However, they were not excluded on the basis of the presence of such problems. As a good substitute for an IQ index, achievement scores were used because they are routinely used as criteria to validate IQ tests (Kaufman, 1990). A score for each child was recorded from either a national or local achievement test. Up to four scores, that is, verbal, quantitative, analytical, and composite scores were all obtained when available, but in all cases at least three scores were obtained, except one case where one score was used. All participants were required to pass a pure-tone, hearing screening (i.e., 20 dB HL for octave frequencies 1000 to 4000 Hz) and also to pass a tympanometry screening (i.e., type A with $\geq$0.2 ml compliance) to ensure normal hearing and normal functioning of the eardrum and middle ear system.

Ninety-three children participated in the assessment process; however, data from 12 participants were eventually excluded because of a diagnosis of mild mental retardation recorded in their school charts, failure on the pure-tone or tympanometry screening, incomplete data, and/or speaking English as a second language. The final data group, therefore, consisted of 81 children (40 boys, 41 girls), age 8 years 8 months to 9 years 9 months.

Prevalence of learning disability (LD), attention deficit hyperactivity disorder (ADHD), and speech/language (SP/L) problems were as follows: 14% (n = 11) were diagnosed with LD, 9% (n = 7) had ADHD, and 4% (n = 3) were receiving SP/L services. Some diagnoses overlapped, however, so a total of 16 children (20% of the sample) was affected by one or more of these problems.

Materials and Procedures

The four tests in the diagnostic battery, the SCAN and the Auditory Fusion Test-Revised (AFT-R) (McCroskey

& Keith, 1996), and instructions were recorded for us on compact disc (CD) by Auditec of St. Louis. All test contingencies, including relative presentation level and monaural/binaural presentation, were accounted for in the CD recording. The four-test battery included the SAAT, Pitch Patterns (PP), Dichotic Digits (DD), and Competing Sentences (CS). SAAT was used in place of the Pediatric Speech Intelligibility (PSI) test originally recommended. Chermak (1996) has noted elsewhere the comparability of the SAAT and PSI, and after consultation with both Chermak (personal communication, 1995) and Musiek (personal communication, 1995), a monaural version (mSAAT) was used for purposes of the present study in place of the PSI because SAAT was judged to have a "user-friendly" advantage over PSI. The SAAT (Cherry, 1980) assesses low redundancy listening (i.e., auditory closure like the PSI), and it requires individuals to listen for a primary stimulus that is embedded in a competing auditory stimulus. The PP test is a test of auditory pattern recognition (involving temporal processing as well) in which subjects must identify high/low pitch qualities in a three-tone series (Pinhiero, 1977). Musiek's (1983) DD test is a test of binaural integration in which subjects hear four numbers concurrently (two in each ear) and must repeat the four numbers aloud. The Willeford (1985) CS test is a binaural separation task that assesses the maturation of the CANS by requiring the subject to attend to a stimulus presented at 35 dB HL in one ear and ignore a different stimulus presented simultaneously at 50 dB HL in the opposite ear. Following the pattern established by Jirsa (1992), the SAAT was presented monaurally to each ear rather than in its traditional diotic format. This procedure was followed to meet the requirement for a monaural low redundancy test. All tests were preceded by formal, recorded instructions.

The underlying factor structure of this recommended battery (henceforth called the Multiple Auditory Processing Assessment ([MAPA]; i.e., mSAAT, PP, DD, and CS) was evaluated with factor analysis. In addition, SCAN's screening utility was compared with failure on one or more of these tests. Specifically, we wanted to know how closely failing scores on the SCAN compared with the battery approach in identifying those who have any one of four auditory difficulties associated with CAPD. This comparison might be viewed as sensitivity, but only for general purposes because admittedly there is no gold standard for CAPD assessment, especially in children for whom physiological confirmation of neurodevelopmental disorder or other neurological involvement (e.g., lesion) remains somewhat elusive. Further, SCAN passing scores were compared with those who passed all of the tests in the battery.

Despite our conviction of the need for a multiple test battery, the possibility was explored of using one of the four tests (mSAAT, PP, DD, or CS) as a screening tool for CAPD or indeed in some combination as a screening/diagnostic battery. All of these tests have been previously recommended as possible screening tests for CAPD (Musiek, Gollegly, Lamb, & Lamb, 1990; DD, Musiek, Gollegly, Kibbe, & Verkest-Lenz, 1991; SAAT, Chermak, Styer, & Seikel, 1995, Cherry, 1980; PP and CS, Musiek, Geurkink, & Kietel, 1982). Again the terms "sensitivity" and "specificity" are used advisedly and only for purposes of comparison between each of the four tests or combination of tests compared with the collective diagnosis from all four.

Consent forms outlining the study were sent home to the parents or guardians of the potential participants. They were also asked to complete the Fisher (1980) Auditory Problems Checklist. Teachers were asked to fill out a similar checklist, the Teachers Scale of Auditory Behaviors ([TSAB]; Schow, Simpson, & Deputy, 1983), with regard to each participant as well as for those children who did not participate. Information obtained from the checklists was used to monitor subject factors, such as history and perception of auditory problems. In addition, the teacher's checklist (TSAB) was used to determine whether participation based on parental permission would yield a representative sample. All potential participants were given the opportunity to consent or refuse to participate in testing. Participants were tested individually with a 65-minute test battery, including the pure-tone hearing screening, tympanometry, SCAN, mSAAT, PP, DD, and CS as well as the AFT-R, which Keith (1997) recommended as a companion measure to SCAN. A shortened version of AFT-R suggested by Keith (1997) was used, in which only 1000 and 4000 Hz were tested. Response modes were identical to those recommended in the test instructions for each test, with the exception of mSAAT. Responses to the mSAAT are typically nonverbal, requiring the child to point to a picture of the stimulus word in a closed-set format. However, to maintain the auditory-only aspect of the test battery, oral responses without the use of visual aids were required (Cacace & McFarland, 1995).

Six audiology clinicians (including D.M.D.) administered the battery of tests to the participants. Clinicians adhered to the pure-tone and tympanometry screening protocols established by ASHA (1997). Scoring parameters followed those recommended by Keith (1986) for SCAN, Keith (personal communication, 1997) and McCroskey & Keith (1996) for AFT-R, Cherry (1980) for mSAAT, Willeford & Burleigh (1994) for CS, and Musiek (1983) for DD. The CS test scoring required perfect repetition of the sentence. Any omission or insertion constituted failure on that sentence. The PP test was scored in two different manners: one counting reversals as correct and one counting reversals as incorrect (Musiek, Geurkink, & Kietel, 1982). The MAPA, SCAN, and AFT-R were administered in 48 random orders to control for order effects such as learning or subject fatigue. However, all three SCAN subtests were given in the order stipulated by standard test format (Filtered Words ([FW], Auditory Figure Ground [AFG], Competing Words [CW].

A recently calibrated MA 39 portable audiometer (Maico) and TA-7A Automatic Impedance Meter (Teledyne Avionics) were used to screen each subject's hearing and middle ear status. The research was conducted in quiet rooms provided by each school. All noise levels and

TABLE 1. Rotated oblique factor loadings using principal components extraction for each of the eight subtest scores along with their groupings within the four emerging factors (*n* = 81).

	APTO	MSC	BI	BS
mSAAT-LE	0.33	**0.78**	0.28	0.14
mSAAT-RE	−0.07	**0.91**	0.13	0.04
PP-LE	**0.95**	0.09	0.30	0.33
PP-RE	**0.96**	0.04	0.27	0.28
DD-LE	0.44	0.07	**0.86**	0.20
DD-RE	0.14	0.36	**0.75**	0.04
CS-LE	0.28	−0.27	0.57	**0.69**
CS-RE	0.34	0.22	0.03	**0.89**

Note. For illustrative purposes the highest loading for each variable is in bold numerals. LE, left ear; RE, right ear; mSAAT, monaural Selective Auditory Attention Test; PP, pitch patterns; DD, dichotic digits; CS, competing sentences; APTO, auditory pattern/temporal ordering; MSC, monaural separation/closure; BI, binaural integration; BS, binaural separation.

TABLE 2. Rotated oblique factor loadings using principal components extraction for each of the three SCAN subtest scores along with their groupings within the two emerging factors (*N* = 81).

	MSC	Binaural
AFG	**0.78**	0.35
FW	**0.85**	0.17
CW	0.31	**0.99**

Note. For illustrative purposes the highest loading for each variable is in bold numerals. AFG, SCAN auditory figure ground subtest; FW, SCAN filtered word subtest; CW, SCAN competing word subtest; MSC, monaural separation/closure; Binaural, Binaural Factor-presumably BS.

calibration data were within the acceptable criteria recommended by ASHA (1997). The noise levels ranged from 24 to 40 dB at 500 Hz, 24 to 36 dB at 1000 Hz, 16 to 30 dB at 2000 Hz, and 12 to 18 dB at 4000 Hz. Portable compact disc players (RCA RP-7913, Sony D-2 Discman, and Panasonic SL-S170) with Optimus Nova-44 stereo headphones were used to conduct the CAPD tests.

Reliability

To ensure intra/interobserver reliability, all clinicians received instructions for scoring before the testing sessions in order to minimize potential differences in scoring. Interobserver reliability was evaluated at least once for each clinician during the research period in an item-by-item analysis between two observers. Final tabulation of scores was performed by D.M.D.

Procedural reliability was primarily accounted for in the CD recording of the tests. However, both a loudness balance judgment and in situ probe microphone measures using the Fonix FP40 were used to determine the volume control setting on the CD players necessary to provide 70 to 75 dB SPL, which is a close approximation to the recommended 50 dB SL, 50 dB HL, and MCL presentation levels for the CAPD tests. This volume control level was fixed throughout testing.

Seven participants (2 boys, 5 girls) were randomly selected to undergo two administrations of the test battery in order to determine test/retest reliability. These two sessions were separated by 1 to 2 weeks from the initial testing.

Results

Factor Analysis

Tabachnick and Fidell (1996) suggest that partial support for a hypothesized factor structure may be obtained when an a priori decision about the number of retained factors produces a final solution that adequately fits the data. Given our theory that four factors underlie test scores, we conducted an exploratory factor analysis in which four factors were retained (see Table 1). As a result, the final solution included a factor with an eigenvalue < 1 (i.e., the eigenvalue for the fourth factor was 0.896). Principal components extraction was used. The oblique rotation was chosen because of the presence of some substantial correlations, particularly between ears and to some extent between tests. In any case, the orthogonal rotation yielded similar impressive factor groupings. To replicate the findings of an earlier study that explored the factor structure of SCAN (Schow & Chermak, 1999), the three SCAN subtests were subjected to an exploratory oblique factor analysis. This analysis produced a two-factor solution even though the second factor had an eigenvalue of 0.77. The two factors had groupings (AFG/FW and CW) similar to those found by Schow & Chermak (1999) in their earlier study (Table 2). Again, the orthogonal results were similar. It was found that CW fell out better with binaural separation (BS) and CS than with the other binaural factor (binaural integration [BI] and DD). Therefore, it was presumed that these two factors from

TABLE 3. Rotated oblique factor loadings using principal components extraction for each of the three SCAN subtest scores and each of the four scores from SAAT and CS along with their groupings within the two emerging factors (*n* = 81).

	MSC	BS
SCAN-AFG	**0.68**	0.05
SCAN-FW	**0.55**	0.07
mSAAT-LE	**0.78**	0.29
mSAAT-RE	**0.74**	0.05
CS-LE	−0.12	**0.84**
CS-RE	0.15	**0.68**
SCAN-CW	0.47	**0.75**

Note. For illustrative purposes the highest loading for each variable is in bold numerals. LE, left ear; RE, right ear; mSAAT, monaural Selective Auditory Attention Test; PP, pitch patterns; DD, dichotic digits; CS, competing sentences; MSC, monaural separation/closure; BS, binaural separation.

TABLE 4. Rotated oblique factor loadings using principal components extraction for each of the three SCAN subtest scores and each of the eight scores from the MAPA along with their groupings within the four emerging factors (n = 81).

	APTO	MSC	BI	BS
SCAN-AFG	0.26	**0.58**	−0.09	−0.44
SCAN-FW	0.27	0.27	0.07	**−0.82**
mSAAT-LE	0.28	**0.75**	0.32	−0.23
mSAAT-RE	−0.06	**0.87**	0.16	0.07
PP-LE	**0.88**	0.08	0.37	−0.26
PP-RE	**0.87**	0.04	0.33	−0.21
DD-LE	0.34	0.07	**0.85**	−0.17
DD-RE	0.04	0.32	**0.69**	0.12
CS-LE	0.32	−0.20	**0.65**	0.09
CS-RE	**0.64**	0.30	0.15	0.50
SCAN-CW	0.47	0.39	**0.65**	−0.11

Note. For illustrative purposes the highest loading for each variable is in bold numerals. LE, left ear; RE, right ear; SCAN-AFG, SCAN auditory figure ground subtest; SCAN-FW, SCAN filtered word subtest; mSAAT, monaural Selective Auditory Attention Test; PP, pitch patterns; DD, dichotic digits; CS, competing sentences; APTO, auditory pattern/temporal ordering; MSC, monaural separation/closure; BI, binaural integration; BS, binaural separation; SCAN-CW, SCAN competing word subtest.

SCAN would correspond to the mSAAT and CS tests from the MAPA. In fact, a consistent factor structure was found given these expectations, when the right and left ear tests for mSAAT and CS were analyzed along with AFG, FW, and CW findings (Table 3). A final exploratory factor analysis that included the three SCAN subtests and the eight subtests of the MAPA was also conducted, and the results were generally impressive. However, although the findings were close to expectations, some instability emerged (Table 4). This instability was not unexpected given the close correlations between the right and left ear subtests and across some of the tests within MAPA.

Only 71 of the 81 children could perform the AFT-R task. These measures were judged to be impractical for the purposes of testing children this age. Final factor findings did not include these results.

Means and SDs

Mean raw scores and standard deviations (*SD*s) for SCAN are shown in Table 5, and mean percentages for the individual tests in the MAPA and the respective *SD*s are shown in Table 6. Table 5 also includes Keith's (1986) means and *SD*s. Mean raw scores for SCAN were used only to allow comparison with Keith's original data. All other SCAN data are presented in terms of percentile rank as calculated from the SCAN manual. This scoring allowed consideration of *SD* in terms of percentile rank. Table 7 demonstrates the 1 and 2 *SD* values derived from the means and *SD*s for the four tests in the battery. In addition, these values are compared with previous normative data, which state performance 2 *SD*s below the mean (Musiek, personal communication, 1995). Use of a t test revealed a significant right/left ear difference for PP (p = .0043) and at the .0001 level for the other three tests from the battery. In terms of gender, only the SCAN-CW and DD, left ear, showed any significant difference based on gender and then only at the $p < .05$ level. Though not significant in most cases, girls tended to score slightly higher than boys on the various tests (see Table 5). In contrast, for Keith's (1986) data, boys tended to have slightly higher scores on the subtests.

TABLE 5. Comparison between mean raw scores and *SD*s from the current study for the SCAN subtests (n = 81) and Keith's (1986) data (n = 130) by age and gender.

	8-Year-Old Children				9-Year-Old Children			
	Boys		Girls		Boys		Girls	
	Current (n = 7)	Keith (n = 68)	Current (n = 16)	Keith (n = 68)	Current (n = 33)	Keith (n = 62)	Current (n = 25)	Keith (n = 62)
FW								
Mean	34.0	32.5	33.4	32.8	35.3	34.5	35.5	33.6
SD	2.9	4.5	3.3	3.3	2.3	3.2	2.7	3.5
AFG								
Mean	31.9	32.4	33.6	31.7	33.5	32.6	34.3	33.0
SD	2.2	3.3	2.7	4.0	2.7	3.7	2.2	3.7
CW								
Mean	78.4	74.6	82.6	79.0	80.4	80.9	83.0	79.0
SD	8.0	14.8	5.2	8.2	7.7	7.4	6.2	8.6
COMP								
Mean	144.3	139.6	149.7	143.4	149.2	148.0	152.8	145.6
SD	11.6	19.0	8.4	11.5	9.5	10.9	8.8	12.0

AFG, SCAN auditory figure ground subtest; FW, SCAN filtered word sub-test; CW, SCAN competing word sub-test; COMP, SCAN composite score.

TABLE 6. Mean percentages and *SD*s for the four-test battery: Right ear and left ear scores ($n = 81$).

	mSAAT-LE	mSAAT-RE	PP-LE	PP-RE	DD-LE	DD-RE	CS-LE	CS-RE
Mean	61.09	65.9	81.7	78.0	78.8	92.3	68.1	89.8
SD	11.53	10.26	20.5	21.1	15.5	7.8	23.5	13.4

LE, left ear; RE, right ear; mSAAT, monaural Selective Auditory Attention Test; PP, pitch patterns; DD, dichotic digits; CS, competing sentences.

Reliability

Interobserver reliability measures were taken on seven occasions. Mean reliability was 0.98. Reliability scores between individual observers were also extremely high (0.94 to 1.00).

Test/retest scores for the seven children given the tests on two occasions were summed over all eight scores of the battery, which yielded a composite score. This resulted in a test-retest correlation of 0.75. Compared to the total group, a restricted range was found for the seven individuals that allowed adjustment of the composite correlation to 0.94. When scores for both ears were summed and adjusted for restricted range, the correlations were 0.89 for mSAAT, 0.99 for PP, 0.54 for DD, and 0.57 for CS. In addition, means and *SD*s of the combined test and the retest scores were compared, and general improvement was noted. A *t* test showed a significant difference at the .01 level between test and retest scores ($p = .0016$), indicating some learning effect.

Sensitivity and Specificity

The favorable factor findings consistent with theoretical expectations, the general consistent pattern of the mean data, and the reliability of the findings all supported the idea of using the battery for comparison purposes with SCAN. On the basis of performance 1 *SD* below the mean on at least one test in the MAPA, 39 of the 81 participants showed at-risk performance for CAPD. This 1 *SD* cutoff was judged to yield too high a failure rate, so presumed sensitivity and specificity data based on the criterion of performance 2 *SD* below the mean on one or more tests in the battery were calculated. Different combinations of tests and their sensitivity and specificity based on 2 *SD* also were calculated, and all the sensitivity and specificity findings are in Table 8.

Participants

A teacher's checklist (TSAB) was used to compare the group of participants with the group of children who were not assessed because their parents did not return the parental permission form. The mean score for the assessed group on the teacher checklist was 40.6, whereas the mean for the nonassessed group was 37.3. Use of a *t* test indicated this was not a significant difference ($p = .40327$).

Discussion

Factor Analysis

The four tests in the battery load into four distinct categories on factor analysis as expected (see Table 1). One factor includes both PP scores and is thus most likely an auditory pattern recognition or temporal ordering (APTO) factor that involves temporal processing. Both mSAAT scores fall into another factor, which we called monaural separation/closure (MSC). We assume mSAAT assesses for monaural auditory performance decrements due to degradation or competition. DD scores constitute another factor, which appears to involve auditory performance requiring BI. Finally, CS scores make up the last factor, which appears to involve auditory tasks requiring BS. Thus, it appears dichotic stimuli may factor into two distinct subcategories, which include integration and separation, as indicated by Chermak & Musiek (1997). Two factors could be expected because of the differences between the two tasks. The MAPA used in this study covers at least three of the five test areas proposed by ASHA (1996) and four definite areas consistent with Chermak & Musiek (1997). The only ASHA test categories not covered are those in binaural interaction and in localization/lateralization. Masking level differences are suggested by

TABLE 7. Performance standards (norms) in percentages at 1 and 2 *SD*s for the current study and 2 *SD*s by age for Musiek (personal communication, 1995).

	mSAAT-LE	mSAAT-RE	PP-LE	PP-RE	DD-LE	DD-RE	CS-LE	CS-RE
1 *SD*	50.6	55.6	61.2	56.8	63.3	84.5	44.6	76.3
2 *SD*	40.0	45.4	40.8	35.8	47.8	76.7	21.1	62.9
Musiek (2 SD)								
Age 8 years			40	40	65	75	40	82
Age 9 years			65	65	75	80	75	90

LE, left ear; RE, right ear; mSAAT, monaural Selective Auditory Attention Test; PP, pitch patterns; DD, dichotic digits; CS, competing sentences.

TABLE 8. Number of subjects who pass and fail the screening test and the diagnostic standard along with sensitivity, specificity, false positives, and false negatives when various tests or test combinations are used as the screen compared with failures on one or more of the MAPA subtests used as the diagnostic standard (*n* = 81).

Screeners	P-Sc P-Dx	P-Sc F-Dx	F-Sc P-Dx	F-Sc F-Dx	Sens (%)	Spec (%)	False Positive (%)	False Negative (%)
SCAN	58	11	3	9	45	95	5	55
mSAAT	61	12	0	8	40	100	0	60
PP	61	14	0	6	30	100	0	70
DD	61	14	0	6	30	100	0	70
CS	61	15	0	5	25	100	0	75
mSAAT/PP	61	7	0	13	65	100	0	35
mSAAT/DD	61	7	0	13	65	100	0	35
mSAAT/CS	61	9	0	11	55	100	0	45
PP/DD	61	9	0	11	55	100	0	45
PP/CS	61	11	0	9	45	100	0	55
DD/CS	61	10	0	10	50	100	0	50
mSAAT/PP/DD	61	2	0	18	90	100	0	10
mSAAT/PP/CS	61	5	0	15	75	100	0	25
mSAAT/DD/CS	61	4	0	16	80	100	0	20
PP/DD/CS	61	6	0	14	70	100	0	30

P, pass; F, fail; Sc, screening; Dx, diagnostic; Sens, sensitivity; Spec, specificity; mSAAT, monaural Selective Auditory Attention Test; PP, pitch patterns; DD, dichotic digits; CS, competing sentences.

ASHA (1996) and Chermak & Musiek (1997) for measuring binaural interaction, but the importance of interaction and localization/lateralization and the question of whether they involve one or two processes will require further study. Nevertheless, it appears that the MAPA used here may have potential. With reference to the factor analysis including SCAN (See Table 3), these findings suggest SCAN covers only two of the factors covered by the battery, because mSAAT loads with SCAN′s FW and AFG subtests and CS loads with SCAN′s CW subtest, just as one would expect. This finding on SCAN also replicates a factor analysis reported by Schow & Chermak (1999) and the initial standardization factor findings of Keith (1986).

When SCAN findings were factor-analyzed with all eight subtests from MAPA, four factors again emerged with eigenvalues > 1.0 (Table 4), but although the groupings were nearly the same, they were less than ideal given expectations from the previous exploratory groupings (Tables 1 to 3). CS-LE and SCAN-CW were found with higher weightings in BI rather than as expected with BS. Given the common binaural nature of BI and BS, this change can perhaps be explained. Also, SCAN-FW had a high weighting in BS, along with CS-RE, rather than with MSC as expected, and CS-RE had unexpectedly an even higher rating on APTO. The changes are not easy to explain, except that the addition of SCAN tests produces somewhat higher weightings distributed across all factors and underscores that several subtests are weaker than the others and not strongly linked with just one factor. SCAN-AFG, SCAN-CW, SCAN-FW, and the CS tests are notable in this regard. Although evidence exists to support the presence of four independent auditory processes, significant common processes such as auditory discrimination are thought to be operative and feed into all of them. Although binaural processes such as integration and separation can sometimes be separately identified, they may at times demonstrate a common core identity. Although separate tests may help us explore different processes, it appears that we should not conclude that these processes are totally independent.

One potential reason for the instability across studies involves the data-driven nature of exploratory factor analysis. Essentially, exploratory analyses generate a best-fitting solution, regardless of any theory about how variables should be related. As such, they are especially vulnerable to chance characteristics of data. Subsequent to the present study, a confirmatory factor analysis was undertaken that was based on the four factor solution to determine how well these four factors actually fit the data. The results were encouraging and supportive of the four-factor solution. A full description of these findings is reported in a companion study (Schow, Seikel, Chermak & Berent, 2000).

Means and SDs

Table 5 compares the mean raw scores and *SD* calculated for performance on the SCAN subtests in this study with those reported by Keith (1986). The means and *SD*s are presented by age and gender in order to maintain consistency with Keith's reported data. Without access to Keith's data, it is impossible to use inferential statistics to compare the two sets of data. However, the findings in this study appear to be very similar to Keith's.

Mean right ear performance was higher than mean left ear performance on each of the four tests in the battery, except PP (see Table 6). This right ear superiority is a common finding in children, because it reflects the development of the CANS and the fact that verbal skills are centered in the left hemisphere. Interestingly, left ear performance was higher on PP. The higher left ear scores may be attributed to the fact that this test consists of pitch stimuli, which presumably, have strong ties with the right hemisphere (Musiek & Pinheiro, 1987).

Means and *SD*s for the four tests in the battery were used to determine the values necessary to identify performance 2 *SD* below the mean (see Table 7). The scores at 2 *SD* in this sample differ somewhat on all three tests (PP, DD, and CS) compared with those reported by Musiek (personal communication, 1995). Musiek reports age-specific normative data, whereas the data in this study combine scores from 8- and 9-year-old children; thus, the two sets of data cannot be directly compared. If Musiek's norms had been used in lieu of establishing separate norms for this sample, the prevalence data discussed later would be even higher.

Reliability

Special care was taken to minimize differences in scoring that might adversely affect the results of this study, resulting in interreliability correlation coefficients of 0.94 or better between all observers.

Composite scores were calculated for the MAPA by combining the right/left ear scores and averaging over all four tests. Using these composite scores, test-retest reliability was excellent at 0.94; reliability on individual tests ranged from 0.54 to 0.99. The DD reliability of 0.54 and the CS reliability of 0.57 were only modest and will need to be studied further for this age group. Good test/retest reliability historically has been difficult to establish when assessing for CAPD (Chermak & Musiek, 1997). Very little data are available for comparison. One sketchy report on a broad sample of children from 8 to 13 years reported test-retest correlations ranging from 0.33 to 0.86 (Hurley, 1990). These correlations included DD, PP, and five other CAPD tests, but no explanation was offered as to how these scores were calculated or on how many individuals. No monaural SAAT test-retest data were available from other sources and even the traditional SAAT manual lists no findings of this type (Cherry, 1980).

With reference to the means and standard deviations for the entire sample, in almost every instance, the comparable means of the test/retest group were higher than the sample as a whole and the standard deviations were lower. The differences from test to retest were significant ($p < .0016$) as shown by a t test. This finding may indicate some variability in performance due to previous exposure to the task. Retests may yield higher scores and be problematic. This concern about retest has also been noted by Amos and Humes (1998).

Sensitivity, Specificity and Prevalence

Keith (1986) recommends consideration of scores 1 *SD* below the mean on the SCAN composite. However, he also indicates it is beneficial to evaluate performance on each subtest to ensure consistency. Because of numerous inconsistencies between subtest scores and composite score, the criterion of 1 *SD* below the mean on one or more of the subtests of SCAN was chosen as at-risk performance. Prevalence of slightly low (1 *SD*) performance on SCAN was found in 22% of individuals (18 of 81). Failure at 2 *SD* on one or more SCAN subtests was found in 14% (11 of 81).

According to ASHA (1996), low performance in even one area (test) indicates the child has an auditory processing concern. The high presumed failure rates (nearly 50%) using a criterion of 1 *SD* below the mean on any one test in the MAPA suggests that this standard overidentifies children at-risk for CAPD. It, therefore, seems prudent to use a criterion in which performance of 2 *SD* below the mean on a single one of the four tests in the battery is suggestive of an auditory processing problem (see Table 7). Using this criterion, 20 participants would be classified as having CAPD, on the basis of the MAPA. Although this failure rate is still high, we chose not to move to 3 *SD*, because 2 *SD* appears to be a good level to examine failure, considering both the MAPA and SCAN. With SCAN 1 *SD* is considered noteworthy by Keith, 1986, and in one case on MAPA, 3 *SD* is below a 0% score, making calculation impractical. Using 2 *SD*s as shown in Table 8 and considering single MAPA tests relative to the battery of tests, sensitivity is highest for mSAAT (40%) and is ≤30% for the other three tests. Specificity on the other hand was 100% in all cases, but this high score was expected because of the criterion used. When individuals pass all the tests and are true negatives, they will also necessarily pass any single test, so any single test will always have 100% specificity. SCAN, also, has very good specificity (95%) but sensitivity of 45%, only slightly better than mSAAT. SCAN and mSAAT are roughly equal as screeners, though neither is acceptable (see Table 8).

Combinations of tests were also considered to determine if they might yield better sensitivity and specificity (Table 8). In other words, would administration of two or three of the tests in the battery constitute an acceptable screening measure? With the fence at 2 *SD*s below the mean on any single test, poor performance on mSAAT in combination with either PP or DD provides the most sensitivity (65%), whereas CS in combination with PP provides the least sensitivity (45%). A combination of mSAAT/PP/DD improved sensitivity to 90%, whereas the worst combination was PP/DD/CS (70%). Specificity for all combinations of tests was again inflated to 100% because of the fact that passing all tests will necessarily mean any one was passed.

Participant Characteristics and Prevalence Adjustment

Because CAPD is generally diagnosed only in the presence of normal intelligence, either a national or local achievement test was used as an indirect measure of intelligence. Three of the children considered to have CAPD demonstrated all of their reported achievement scores at or below the 16th percentile (<1 *SD*). These are the children who should probably be labeled as slow learners, precluding a designation of pure CAPD. Jirsa (personal communication, 1997) indicates it is nearly impossible to make a diagnosis of CAPD in the presence of low IQ scores because it is difficult to separate whether the problems are associated with auditory processing or with higher functioning. Thus, of the 20 children who performed below 2 *SD*s on one or more of the four auditory tests, three were considered slow learners. Of the remaining 17 who appeared to have serious auditory difficulty (≤2 *SD*s), two were diagnosed with LD, two were identified with ADHD, one had both LD and ADHD, and one was receiving SP/L services. These children should not be excluded from possible CAPD even though they have other problems. The number of children ($n = 16$) in the sample of 81 with these special conditions who also seemed to have serious auditory problems was about one third (LD = 3/11 = 27%, ADHD = 3/7 = 43%, SP/L = 1/3 = 33%). This linkage with other disorders indicates convergent validity, because the children with these disorders demonstrate substantial prevalence of auditory difficulty. This comorbidity of CAPD, ADHD, and LD has been cited by a number of authors (Chermak & Musiek, 1997; Keith & Engineer, 1991; Riccio, Hynd, Cohen, & Gonzales, 1993).

To summarize, using the criterion of performance 2 *SD*s below the mean on at least one test in the MAPA, after excluding slow learners, the prevalence of presumed CAPD in this sample is 21%, or 17 of 81 children. In comparison to a predicted prevalence of 2 to 3% (Chermak & Musiek, 1997), 21% seems high, but that report was based on labeling CAPD in the case of performance 3, rather than 2, *SD*s below the mean. Furthermore, several other factors may bring a closer convergence on prevalence estimates.

Consistently low performance on multiple test measures within a factor area only occurred about one half of the time in this study when MAPA and SCAN scores were both reviewed. Lack of consistency of poor scores may reduce the number of failures further if the children were given additional tests within an area before confirming a diagnoses.

In addition, prevalence may also be adjusted by looking at inconsistencies between performance on the CAPD battery and some other criteria, such as teacher/parental report or performance on achievement tests. Such analysis was beyond the scope of this study, but could be reported in future work. It also seems possible that 2 to 3% may be too low a prediction, and actual prevalence may be higher when there is an accepted gold standard.

Procedural Issues

Historically, the PP test has been scored in at least two different manners. In one, reversals of pitch identification but maintenance of the pattern (e.g., high-low-high is presented to the participant but is recounted as low-high-low) are counted as correct, whereas, in the other, reversals are counted as incorrect. Factor analysis indicated both manners of scoring load into the same area, thus suggesting the manner of scoring does not significantly change the process being assessed, though it may slightly affect whether or not a child will be identified as at-risk. We allowed credit for reversals in our final tabulations.

Implications and Summary

One of the most serious challenges in CAPD research is defining a gold standard for diagnosis. We used a variation of the test battery proposed by Musiek & Chermak (1994) as a tentative standard. On the basis of the present factor findings, the four tests that we used appear to cover at least three ASHA (1996) test areas and one of these areas in terms of two distinguishable subareas. Because ASHA (1996) proposed five test areas, there may be two areas not covered by the four tests (i.e., binaural interaction and/or localization/lateralization). Further study and factor analysis will need to be performed to resolve the binaural interaction and localization/lateralization questions and to assess exactly how other CAPD tests load in comparison to the MAPA in this study. Continued research in this area may either verify the ASHA document or change the definition of what factors contribute to an auditory processing disorder. For now, MAPA may provide an appropriate set of multiple tests that approximates the recommendations of the ASHA consensus report and, as proposed by Musiek and Chermak (1994), tests relevant functional areas within the CANS.

One purpose of this article was to consider if there might be a reliable and valid screening protocol for CAPD using for confirmation a reasonable battery of tests based on the recent ASHA (1996) recommendations. It appears that none of the measures used alone constitutes a very good screen compared with findings from a multiple test battery. SCAN does not have high sensitivity for CAPD in relation to the MAPA used in this study, and preliminary findings suggest the three SCAN subtests cannot be helped much by the addition of AFT-R. AFT-R could not be completed by a number of children in a shortened form and is probably too long in its entirety to use as part of a screening procedure. No single measure within the MAPA works very well as a screen. Sensitivity only exceeds 50% when using a combination of two or three of the tests. A three-test screen may be suitable because sensitivity actually reaches a desired 90% in one instance. However, all four tests together only require about 30 min to administer, not much longer than the SCAN (20 min) alone or the best three-test combination. Thus, all four tests could be used for initial screening or preliminary diagnostic testing, with more tests being crucial to confirm the diagnosis.

Additional tests beyond the MAPA may be selected in the future to provide increased precision and round out a diagnostic battery that could follow-up on the initial screening. After the MAPA, SCAN could be used to sample and confirm poor performance in perhaps two of the test areas. For example, SCAN′s FW and AFG subtests could sample further the area assessed by mSAAT. However, other tests that cluster better with MAPA subtests in factor studies need to be developed to confirm poor results in all four areas. Perhaps a diagnosis would not be warranted if consistently low performance was not noted with more than one test that purportedly measures the same content area. This type of follow-up testing in at-risk areas might thus reduce the high prevalence noted above. Further studies on reliability also are needed, but reliability of assessment within an area might simply be improved by more extensive sampling of an area with two or three tests.

In short, use of the MAPA is encouraging but is still in a preliminary stage. Factor findings suggest it has an apparent utility as a multiple-test CAPD battery for third-grade children, and it meets some of the objectives described in the consensus document proposed by ASHA (1996). Further study will be needed to determine the ultimate utility of this battery in the diagnosis of CAPD. If we are able to classify our diagnostic efforts in a more precise way through such a battery, this precision will be promising because it will facilitate the remediation of these children and help us more effectively evaluate the results of those therapy efforts.

Acknowledgments

This work was originally completed by the first author in a thesis as part of a master's degree program at Idaho State University, with the second author serving as major advisor. Gail Chermak has been extremely generous with her support, insight, and encouragement from the inception to the completion of the work. Frank Musiek, Jack Willeford, and Bill Carver granted permission for the use of tests for which they hold exclusive rights. Jeff Brockett and Bill Carver helped perfect the CD recording used herein. Thayne Smedley, Tony Seikel, and Mark Roberts offered helpful advice and input. Mary Whitaker, school audiologist, the Pocatello School District administration, three elementary schools, and their teachers and staffs were generous in helping us locate the willing children who gave their time and attention to this project. Matt Berent and Teri Peterson rendered invaluable assistance with the statistics used herein. To everyone who helped so generously we offer our grateful, heartfelt thanks.

References

American Speech-Language-Hearing Association. (1996). Central auditory processing: Current status of research and implications for clinical practice. *American Journal of Audiology, 5*(2), 41–54.

American Speech-Language-Hearing Association. (1997). Committee on Audiometric Evaluation. *Guidelines for identification audiometry* (p. 40).

Amos, N. E., & Humes, L. E. (1998). SCAN test-reliability for first- and third-grade children. *Journal of Speech, Language, and Hearing Research, 41,* 834–845.

Cacace, G. D., & McFarland, D. (1995). Modality specificity as a criterion for diagnosing central auditory processing disorders. *The American Journal of Audiology, 4,* 36–48.

Chermak, G. D. (1996). Central testing. In S. A. Gerber (Ed.), *Handbook of pediatric audiology* (pp. 206–253). Washington, DC: Gallaudet University Press.

Chermak, G. D., & Musiek, F. E. (1997). *Central auditory processing disorders: New perspectives.* San Diego: Singular.

Chermak, G. D., Styer, S. A., & Seikel J. A. (1995). Study compares screening tests of central auditory processing. *Hearing Journal, 48*(5), 29–34.

Cherry, R. S. (1980). *Selective auditory attention test*. St Louis: Auditec.

Fisher, L. (1980). *Fisher's auditory problems checklist.* Cedar Rapids, IA: Grant Wood.

Hurley, R. M. (1990). Decision matrix analysis of selected children's central auditory processing tests. *Journal of the American Academy of Audiology, 1*(1), 50.

Jerger, S., Johnson, K., & Loiselle, L. (1988). Pediatric central auditory dysfunction: Comparison of children with confirmed lesions versus suspected processing disorders. *American Journal of Otology, 9,* 63–71.

Jirsa, R. E. (1992). The clinical utility of the P3 AERP in children with auditory processing disorders. *Journal of Speech and Hearing Research, 35,* 903–912.

Kaufman, A. S. (1990). *Assessing adolescent and adult intelligence* (pp. 19–20). Boston: Allyn and Bacon, Inc.

Keith, R. W. (1986). *SCAN: A screening test for auditory processing disorders*. 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 American Academy of Audiology, 6*(4), 286–292.

Keith, R. W. (2000). *SCAN C test for auditory processing disorders in children—revised.* San Antonio: The Psychological Corporation.

Keith, R. W., & Engineer, P. (1991). Effects of methylphenidate on the auditory processing abilities of children with attention deficit-hyperactivity disorder. *Journal of Learning Disabilities, 24,* 630–636.

McCroskey, R. L., & Keith, R. W. (1996). *Auditory fusion test—revised: Instruction and user's manual*. St. Louis: Auditec.

Musiek, F. E. (1983). Assessment of central auditory asymmetry; the dichotic digit test revisited. *Ear and Hearing, 4*(2), 79–83.

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 of assessment of auditory perceptual dysfunction in children. *Laryngoscope, 92,* 251–257.

Musiek, F. E., Gollegly, K. M., Kibbe, K. S., & Verkest-Lenz, S. B. (1991). Proposed screening test for central auditory disorders: Follow-up on the Dichotic Digits test. *American Journal of Otology, 12,* 109–113.

Musiek, F. E., Gollegly, K. M., Lamb, L. E., & Lamb, P. (1990). Selected issues in screening for central auditory processing dysfunction. *Seminars in Hearing, 11,* 372–384.

Musiek, F. E., & Lamb, L. (1994). Central auditory assessment: An overview. In Katz, J. (Ed.), *Handbook of clinical audiology* (4th ed., pp. 197–211). Baltimore: Williams & Wilkins.

Musiek, F. E., & Pinheiro, M. L. (1987). Frequency patterns in cochlear, brainstem, and cerebral lesions. *Audiology, 26,* 79–88.

Pinheiro, M. L. (1977). Test of central auditory function in children with learning disabilities. In R. Keith (Ed.), *Central auditory dysfunction* (pp. 223–256). New York: Grune & Stratton.

Riccio, C. A., Hynd, G. W., Cohen, J. J., & Gonzalez, J. J. (1993). Neurological basis of attention deficit hyperactivity disorder. *Exceptional Children, 60*(2), 118–124.

Schow, R., & Chermak, G. D. (1999). Implications from factor analysis for central auditory processing disorders. *American Journal of Audiology, 8*(2), 137–142.

Schow, R., Seikel, J. A., Chermak, G. D., & Berent, M. (2000). Central auditory processes and test measures: ASHA 1996 revisited. *American Journal of Audiology*, 9, October 5.

Schow, R., Simpson, J., & Deputy, P. (1983). Comparison of two behavioral screening scales for auditory processing disorders. ASHA Western Regional Conference, Honolulu.

Singer, J., Hurley, R. M., & Preece, J. P. (1998). Effectiveness of central auditory processing tests with children. *American Journal of Audiology, 7,* 73–84.

Tabachnick, B. G., & Fidell, L. S. (1996). *Using multivariate statistics* (3rd ed.). New York: Harper Collins College Publishers.

Willeford, J. A. (1985). Assessment of central auditory disorders in children. In M. L. Pinheiro & F. E. Musiek (Eds.), *Assessment of central auditory dysfunction: Foundations and clinical correlates* (pp. 239–250). Los Angeles: Williams & Wilkins.

Willeford, J. A., & Burleigh, J. M. (1994). Sentence procedures in central testing. In J. Katz (Ed.), *Handbook of clinical audiology* (4th ed., pp. 256–268). Baltimore: Williams & Wilkins.

Received October 25, 1999
Accepted June 1, 2000
First published (online) October 5, 2000
http://journals.asha.org
DOI: 10.1044/1059–0889(2000/012)

Contact author: Ron Schow, Idaho State University, Pocatello, ID 83209. Email: Schorona@isu.edu