FORM EQUIVALENCY ON THE BETA III VERSION OF MULTIPLE AUDITORY PROCESSING ASSESSMENT (MAPA)

by

Laurie Conlin

A paper

submitted in partial fulfillment

of the requirements for the degree of

Master of Science in the Department of Audiology

Idaho State University

August 2003

Acknowledgments

A special thanks goes to my thesis advisor, Dr. Ron Schow, for having so much enthusiasm for this project. He was very encouraging and always willing to help in whatever way he could.

Many thanks also goes to Dr. Tony Seikel and Dr. Mark Roberts for serving on my thesis committee, answering my questions, and providing willing help. Dr. Loren Toussaint and Teri Peterson were also wonderful to help with the statistics and data analysis.

Thanks to Jenn Holst and Vicki Anderst, school district audiologists, for their encouragement and assistance. Also, many thanks goes to the wonderful teachers, principals, and children who participated in this study. They were a delight to me.

My professors, classmates, and friends were also wonderful to help when possible and provide moral support when I was stressed. Thank you all.

An enormous thank you goes to Sherry Summers, my companion study partner. I could not have completed this project without her help, support, and sense of humor.

Finally, I want to thank my family. They offered continued support, encouragement, and love throughout the entire project. They also reminded me of my many blessings, for which I must thank the Lord.

CONTENTS

LIST OF TABLES

Table

Abstract

The American Speech Language Hearing Association (ASHA), has suggested guidelines for evaluating auditory processing disorders (ASHA, 1996), but currently there is no standard testing tool that meets all of the guidelines. In 1997, Domitz and Schow proposed a test battery named the Multiple Auditory Processing Assessment (MAPA) and provided data from the Beta I version of it which attempted to incorporate most of these guidelines. Shiffman, 1999, gathered more data from the Beta II version of MAPA, but more refinement was necessary.

The Beta III MAPA is now more difficult and includes information for a form A and B. These forms were tested on school children ages 8-11 for equivalence and to see if the ceiling effect had been eliminated while still maintaining reasonable test results for children 8-11.

Means, standard deviations, Pearson correlations, fmax, and t-tests were used to examine the subtests of MAPA: monaural Selective Auditory Attention Test (mSAAT), Dichotic digits (DD), Competing Sentences (CS), Pitch Pattern (PP), and two new subtests, Speech-in-noise for Children and Adults (SINCA) and Duration Pattern (DP). Grouped monaural, binaural, and temporal tasks and a composite score for the entire test were also evaluated.

Monaural: When both ear scores are combined on each mSAAT and SINCA, forms A and B are moderately related (r=.37 to .40) and the means are not significantly different (p>0.01). When msAAT and SINCA are combined, the correlation coefficient r=.46. Thus, these forms are passably acceptable for

equivalence at least until SINCA is modified for some needed improvement.

Binaural: CS and DD, with r's between .69-.78 for forms A and B show acceptable equivalency. However, significant mean differences on DD between forms suggest a learning effect. CS and DD combined into one binaural score yields a correlation of r=.81.

Temporal: PP and DP, show excellent correlations between forms, r=.89 and .85, respectively. However, DP means for forms A and B are different (p=0.002). For this and other reasons, DP may not be retained in MAPA. The new test, Quick Tap (from a companion study), and PP scores were combined with an r=.90 and no difference between mean scores (p=0.727).

In summary, all tests on both forms demonstrate at least some measures of equivalency. The ceiling effect has been eliminated for all tests, and test results appear reasonable in most respects as compared to results on Beta I and Beta II MAPA.

CHAPTER I

Introduction

The process of screening and identifying children with Auditory Processing Disorders (APDs), previously called Central Auditory Processing Disorders (CAPDs), continues to be a debated topic in the field of audiology. According to Hall (1999), “the term CAPD is used to describe a deficit in the perception or complete analysis of auditory information due to central auditory nervous system dysfunction, usually at the level of the cerebral cortex” (p. 35). Jerger and Musiek (2000) state that this deficit in information processing is specific to the auditory modality. The children diagnosed with APD have normal peripheral hearing. However, they are often unable to process certain aspects of auditory information correctly. Much of the current debate centers on which tests are appropriate and should be used in the screening and diagnosis of APD.

Many tests and procedures have been recommended for screening and diagnosing APD. According to the 1996 consensus statement by the American Speech-Language-Hearing

Association (ASHA), the central auditory processing system is responsible for six behavioral processes. These six processes are sound localization and lateralization, auditory discrimination, auditory pattern recognition, temporal aspects of audition, auditory performance decrements with competing acoustic signals, and auditory performance decrements with degraded acoustic signals. To be diagnosed with APD, a child must exhibit a deficit in one or more of these processes (ASHA, 1996). ASHA recommended five behavioral auditory test measures for these six areas. As Schow, Seikel, Chermak, and Berent (2000) point out in a follow-up article, there is not a definite correspondence between all six behavioral processes and the five recommended testing measures of ASHA. They recommended behavioral tests that included a monaural task, a pattern or temporal ordering task, and two binaural tasks that involve integration and separation. These four tasks are thought to relate to three important behavioral processing areas (monaural tasks, binaural tasks, and pattern recognition tasks).

Jerger and Musiek (2000) reported on a consensus conference (Bruton) of 14 audiologists that discussed, among other APD issues, the diagnosis of APD for school children. This conference recommended a minimal behavioral test battery as well as the use of electro/physiological/acoustic tests and neuroimaging studies. Chermak (2001), who participated in the Bruton conference, summarized the recommendations there and suggested that the behavioral test battery should contain at least one test in three areas: 1) temporal processing, 2) binaural processing, and 3) monaural low-redundancy speech recognition. In addition, she recommended that although APD is diagnosed by audiologists after an extensive evaluation, a “comprehensive evaluation requires a multi-disciplinary team approach” (p. 12). She recommended the inclusion of speech-language pathologists, psychologists, and educators in the evaluation process.

Katz et al. (2002) supported the behavioral tests of the Bruton group but disagreed with the use of electro/physiological/acoustic tests and neuroimaging for the general population because they felt that these measures were unrealistically expensive and time consuming. They also stated that research does not confirm that children with APD are significantly different in electro/physiological/acoustic measures from children who do not have APD. In addition, little is known about imaging. In response, Jerger and Musiek (2002) stated that, “if we are ever going to have a gold standard for APD, it will probably be in the form of electrophysiological measures” (p. 20). Very little has been done so far on electrophysiological measures of APD while most work has focused on behavioral tests.

In an attempt to provide a reliable behavioral test battery through the use of factor analysis, Domitz and Schow (2000), administered a battery of APD tests to school-aged children. They named their test battery the Multiple Auditory Processing Assessment (MAPA). While they felt the MAPA was an appropriate test, tapping into several important areas from the 1996 ASHA guidelines, the ceiling effect was occurring with some of the tests and especially with older children.

Recently, a revised Beta III version of the MAPA has been developed and recorded on compact disc. The Beta III MAPA includes test information for a Form A and Form B version of the test. The Beta III MAPA has all of the same tests as the original MAPA plus four more. These behavioral tests are organized into three areas (monaural, binaural, and temporal patterns) as recommended by Chermak and (monotic, diotic, and dichotic) as advised by the Bruton group. In addition, the tests that were prone to the ceiling effect were modified so the tasks would be more difficult. The specific modifications to the tests will be described in the literature review (Appendix A).

Research Questions

There were two research questions asked in this study.

1. Are form A and form B of the four original tests within the Beta III MAPA (mSAAT, PP, DD, CS) and two of the experimental tests (SINCA, DP) equivalent tests based on t-tests, fmax, and Pearson correlations?
2. Are the four modified tests appropriately designed to overcome the ceiling effects and produce reasonable means and standard deviations?

CHAPTER II

Methodology

The purpose of this study was to determine if both A and B forms of the Beta III Multiple Auditory Processing Assessment (MAPA) tests are equivalent. Improving the revised MAPA will allow audiologists to evaluate three important behavioral test areas as described in the 1996 ASHA guidelines and at the Bruton Conference.

Procedures

Participants were obtained by working with the Blackfoot/Snake River Idaho school district audiologist. Initially, all students in the selected classrooms received a consent form to be taken home to the parents explaining the purpose of the study (Appendix C). In addition, parents were asked to complete a twelve-item checklist based on the work of Shiffman (1999) and Chermack, Somers, and Seikel (1998) to give information on their child’s attention and auditory behavior (Appendix D). Teachers were also asked to complete the same scale for each child participating in the study as a verification measure for the parental report.

All children who returned the consent form were told that they could refuse participation later, despite parental consent. All children, after completion of the test, were asked if they would be willing to be retested. All testing was administered by graduate-level clinicians, certified audiologists, or trained assistants according to ASHA guidelines and the guidelines for scoring the original MAPA. The children returning the consent forms who agreed to participate were tested in two testing sessions.

During the first test session, the children received a hearing screening consisting of pure tones, tympanometry, and OAEs. To pass the hearing screening, children had thresholds of 20 dB HL or better at 1000, 2000, and 4000 Hz. In addition, immittance testing needed to show compliance greater than .2 ml. If this was not shown, the child was required to pass the hearing screening at 250 and 500 Hz.

A calibrated Maico (MA39) portable audiometer with TDH39 headphones, an Earscan immittance screener, and a portable Maico Ero-scan otoacoustic emissions screener were used to screen each subject’s hearing and middle ear status. All equipment was calibrated following ANSI guidelines. Daily calibrations on the tympanometer were conducted as well as biological listening checks on the audiometer and CD players.

Portable Lenoxx Sound Model CD-87 compact digital audio disc players with digital Koss (UR15) or Optimus Nova-44 stereo headphones were used during APD testing. The administering clinicians used monitoring earbuds or headphones while conducting the testing. The auditory processing test was delivered to the subject at an approximate level of 50 dB HL. Following the procedures of Domitz (1997) and Shiffman (1999), and to ensure delivery consistency, the volume control of the CD player was fixed at 75 dB SPL throughout testing to approximate a 50-55 dB HL presentation level as recommended. Since the monitoring headphones/earbuds were used, this level was established with the testing headphones and the additional set of phones attached to the CD player through a Y-cord.

After the initial hearing screening, each participant was tested individually for a 30-minute test battery, using form A. At the end of the first session, the children received McDonald’s coupons. The following six tests, shown in the order they are recorded on the CD, were administered. The tests were administered in the order recorded on the CD. However, the starting order of administration varied between the children to control for threats of validity involving subject fatigue. This procedure was followed for both forms A and B (first and second sessions).

3. Monaural selective auditory attention test (mSAAT)
4. Pitch patterns (PP)
5. Dichotic digits (DD)
6. Competing sentences (CS)
7. Duration patterns (DP)
8. Speech in Noise for Children and Adults (SINCA)

Since this data was going to be collapsed with a companion study, all the children were given two other experimental tests, a Gap Detection task and the Quick Tap (described in a companion study), as part of the form A testing procedure.

The second session was completed during the following week (elapsed time was 7-10 days between tests) with the children being tested individually for 30 minutes using the same six tests but from form B. About half of the sample (25 children) were also given the experimental Quick Tap and Gap Detection tasks. These two tasks do not have multiple forms. The reward for the second session was either more McDonald’s coupons or a selection of stickers or small treats.

All instructions for the tests were pre-recorded on the CD. Therefore, the clinicians were responsible for monitoring the CD player through the additional set of headphones/earbuds, ensuring correct placement of the testing headphones, clarifying instructions when requested by the child, and scoring each response as correct or incorrect. The CD player was paused only during school disruptions, breaks for the children, and for the instruction clarification. It was not paused to allow more time for the children to respond.

The answer sheets were coded to ensure confidentiality. To ensure measurement reliability, the researcher trained the clinicians prior to beginning the testing and observed each clinician at least once during the research period. The researcher also double checked all final score tabulations to ensure accuracy.

Instrumentation

The instrument used included the original four tests from the Beta III MAPA as revised, using both forms A and B. Four additional experimental tests were also included in the test battery with two evaluated for equivalency in this study. The MAPA will eventually consist of tests in each of the areas of monaural, binaural, and temporal processing tasks. The determination of the tests to be included depends on the results of this study and a companion study. The monaural tests were the monaural Selective Auditory Attention Test (mSAAT), and the Speech in Noise for Children and Adults (SINCA). The binaural tests were the dichotic digits (DD), and competing sentences (CS). The temporal tests included pitch patterns (PP) and duration patterns (DP). The results from two experimental tests, forms A and B of the SINCA and DP, a monaural and a temporal task, are included in this study. The Quick Tap test and the Gap Detection test were used also but evaluated in the other study. The tests are on a compact disc (CD) recorded by Auditec, a major supplier and developer of auditory tests in St. Louis, Missouri. All tests are preceded by formal, recorded instructions and coincide with the answer sheets. The administering clinicians recorded each test based on the subject response or non-response. Answer sheets for forms A and B appear in Appendix B.

CHAPTER III

Results and Discussion

To review, there were two research questions asked in this study.

9. Are form A and form B of the four original tests within the Beta III MAPA (mSAAT, PP, DD, CS) and two of the experimental tests (SINCA, DP) equivalent tests based on t-tests, fmax, and Pearson correlations?
10. Are the four modified tests appropriately designed to overcome the ceiling effects and produce reasonable means and standard deviations?

This section discusses the noise level measures and participants. Mean scores, standard deviations, Pearson correlations, fmax, and paired t-test results were used to examine the equivalency between forms and examine any ceiling effects. In addition, a discussion of the mean scores, standard deviations, Pearson correlations, fmax, and t-test results are given when tests are examined in combined groups of monaural, binaural, and temporal tasks.

Noise Levels

Testing was completed in a quiet room provided by the school. The rooms met the noise-level requirements recommended by ASHA for the 500-4000 Hz range which is that noise should be less than 46 dB at 500 Hz, 49.5 dB at 1 kHz, 54.5 dB at 2 kHz, and 62 dB at 4 kHz (ASHA, 1997). Ambient noise was monitored using the Quest-188 sound level meter and rechecked when noise levels changed noticeably. At no time were measurements taken that exceeded the ASHA guidelines. Table 1 shows the recorded noise measurements for each school compared to the 1997 ASHA guidelines.

Table 1. Recorded noise measurements for each school

Hz	ASHA	School 1	School 2
500	46 dB	33.6 dB	39.0 dB
1000	49.5 dB	25.9 dB	32.5 dB
2000	54.5 dB	21.6 dB	27.0 dB
4000	62 dB	19.2 dB	20.6 dB

Participants

Parent permission forms (see Appendix C) and auditory behavior scales (Appendix D) were delivered to an Idaho elementary school in both the Blackfoot and Snake River school districts. Principals and teachers of the third and fifth grade classrooms were contacted in advance and had agreed to participate in the study (see Appendix E). The children were informed about the testing and were asked to have their parents sign the permission slips and fill out the questionnaires. Demographic information gathered from the parent, child, teacher, and school records included age, gender, handedness, and whether the child had been diagnosed with either attention-deficit hyperactivity disorder, speech or language problems, learning disabilities, IQ deficit or difficulties with math, reading, or writing.

Fifty children returned forms and volunteered to be subjects. Of these, two were eliminated after the data collection based on cognitive impairment and age considerations. The results for forty-eight children are included in this study. There were 23 participants from one school and 25 participants from the other. Twenty-five children in the third grade participated in the study and 23 children from the fifth grade were involved. There were 29 females and 19 males. Four of the children were left-handed. Table 2 shows a summary of participants in the study by school and grade.

Table 2. Summary of participants

	3rd grade	5th grade	Total
School 1	23	- - -	23
School 2	2	23	25
Total	25	23	48

The subjects came from five third grade classrooms and three fifth grade classrooms and represented a diverse socioeconomic status. The majority were Caucasian; however, several children, not Caucasian, judged by their teachers to speak English with proficiency (native or near-native ability), were also included in the study. A division by age showed nine eight-year-olds, 16 nine-year-olds, 13 ten-year-olds, and ten eleven-year olds. All of the third graders were eight to nine years of age. The fifth graders included in the study were 10-11 years of age. All of the results are examined and shown by grade level and thus grouped in two age groups (8-9, 10-11).

Participation was dependent on the return of the parental consent form and the passing of a pure-tone hearing screening for both ears. An immittance and electrophysiological screening (Otoacoustic emissions or OAE) was performed on each child to gather information that might clarify electroacoustic and APD relationships. The absence of emissions alone did not eliminate subjects. Two of the children were found to have “refers” in the left ear on OAE. Both also showed type B tympanograms. However, both passed the hearing screening in the left ear which included 250 and 500 Hz. Therefore, no children were eliminated from the test due to the hearing screening.

All children in all classrooms where the principals and teachers agreed to participate were given the parental consent form to have completed and return. Children were included in this phase of the study randomly and as part of a larger data gathering on the MAPA. All children in each of the classrooms who qualified were tested. No special consideration in choosing participants was given to factors such as academic performance or teacher recommendation; however, children were excluded due to a known diagnosis of mental retardation or limited English proficiency. These data were gathered after testing was completed. One child was found to have a cognitive impairment (IQ < 70). In addition, there were two children who fell outside the 8-11 age range selected for this study. One of these children was the child with the cognitive impairment. Thus, these two test results were not included in the study and the sample size was dropped to n=48.

Children with a diagnosis of ADHD would not have been excluded from this study since they are part of the general population for which the MAPA may be used to screen for APD. None of the children in this sample had this diagnosis. Three children had a diagnosis of learning disorders. Of these three, one child received speech and language services. The other two children both received additional services as part of their Individual Education Plan (IEP). Both received services for reading and writing, and one student for math.

Question 1

Statistical analysis of means and standard deviations

Means and standard deviations for forms A and B of the monaural tasks mSAAT and SINCA are shown in Table 3. The results are shown by ear and by age group. Mean monaural mSAAT scores were generally in the range of 10-12 out of 25 items, or about 40-50%. The mean scores for the fifth graders were slightly higher (1 or 2 items) than for the third graders on both forms as expected. A t-test showed a mean difference of .4 between the left and right ear mean scores ($p=0.381$) for form A. Form B showed a mean difference of .5 between ears ($p=0.313$). Since there was little difference between ears, it seemed reasonable to combine both scores for the left and right ears for one overall monaural score for each grade. On the mSAAT, the overall mean score out of 50 items for third graders was 23.5 for form A and 21.3 for form B. For the fifth graders, the overall score was 24.2 for form A and 23.6 for form B. The fmax test was used to determine if any differences in standard deviations were significant. There were not any significant differences in standard deviations for any mSAAT task ($p>0.05$).

The SINCA shows scores as Signal to Noise Ratios (SNR) with lower scores indicating better performance on the test. There are 24 test items for each ear. A more complete description of the test and the scoring procedure is given in the literature review. In general, the lowest score would be a SNR of 0. Originally, scores were tallied without giving the children any practice items, thus including all 24 items. Initial results showed poorer mean scores on the first test that were not shown on the other tests, suggesting a need for practice items to ensure that the children understood the task. Thus, the first four items on each ear were used as practice items and the scoring was completed on the final 20 items.

The mean scores using the 20 items were between 5-7 for third graders and between 3-6 for fifth graders, which was consistent with the expectation that SNRs would get better (lower) as children get older. A t-test showed a mean ear difference of .1 for form A ($p=0.794$). Since little difference was found between ears, scores were summed for a combined monaural SINCA form A score. For form B, there was a mean ear difference of 1.8, with the left ear being the poorer ear ($p=0.000$). While this mean difference is statistically significant, the scores for form B were also combined in order to be used as a comparison. The third graders showed 5.6 SNR for form A and 5.8 for form B when combined. The fifth graders showed a combined mean score of 3.8 SNR for form A and 4.2 SNR for form B. These findings indicate an improvement of 1-2 dB between the two grades.

For the standard deviations, the fmax ratio results showed significant differences at the .05 level for the combined SINCA score for third graders as well as the right SINCA and combined SINCA for fifth graders. However, these findings were not significant at the .01 level.

Table 3. Means, standard deviations, and fmax ratio for monaural tasks for right, left, and both ears combined
**indicates significance at p<0.05

Grade	Test	Ear	Form A		Form B	fmax ratio
3rd (8-9 years)	mSAAT	R	¯ SD	11.8 2.4	11.3 2.6	1.17
	mSAAT	L	¯ SD	11.6 2.4	10.0 2.9	1.46
n=25	SINCA	R	¯ SD	5.6 2.8	5.1 1.9	2.17
	SINCA	L	¯ SD	5.7 2.4	6.4 2.3	1.09
	mSAAT	Both	¯ SD	23.5 4.0	21.3 4.5	.79
	SINCA	Both	¯ SD	5.6 2.2	5.8 1.3	2.86**
5th (10-11 years)	mSAAT	R	¯ SD	12.4 3.3	11.6 2.5	1.74
	mSAAt	L	¯ SD	11.8 2.5	12.0 2.9	1.35
n=23	SINCA	R	¯ SD	3.8 3.1	3.0 1.9	2.66**
	SINCA	L	¯ SD	3.9 2.1	5.4 1.9	1.22
	mSAAT	Both	¯ SD	24.2 4.6	23.6 4.2	1.20
	SINCA	Both	¯ SD	3.8 2.3	4.2 1.4	2.65**

The binaural tasks (CS and DD) are shown in Table 4.

The highest possible score for the CS task was 30 items correct in each ear for a total score of 60. The CS scores are shown for the individual ears and as a total score for both ears. The third graders demonstrated a mean score in each ear between 11-13 with no apparent systematic differences based on form or ear. A t-test showed that mean differences between ears was not significant for either form A or form B ($p=0.819$ and $p=0.644$ respectively). Again, a monaural combined score was devised which showed a mean score of 23.9 for form A and 22.3 for form B. The fifth graders showed a total mean score of 28.3 for form A and 27.1 for form B. As before, a slight improvement was seen in the mean score for older subjects. The fmax ratios for standard deviations do not show any significant differences on the binaural tasks of CS or DD ($p>0.05$).

For the DD task, two different scoring methods were used. One procedure calculated the total score out of 120 possible items correct, repeated in a free recall condition. The other method looked at scores for the items in the right and left ears correctly repeated by ear as directed. For form A, the third graders scored between 24-26 items correct out of 60 total items for the left and right ears. The fifth graders scored between 27-31 items correct for the left and right ears. A t-test showed the mean difference between ears to be 2.25, which was not statistically significant ($p=0.13$). For form B, the scores ranged between 27-30 for the third graders and 34-38 for the fifth graders. The t-test showed a mean difference of 2.42, or 4%, which was not significant at the .01 level ($p=.041$). Using this scoring procedure, an ear difference is not judged to be clinically relevant.

Neijenjuis, Snik, Priester, van Kordenoordt, and van den Broek (2002) tested children ages 9-12 on a similar dichotics digit task. Their procedure allowed for a free recall condition with scores reported for only the right and only the left ear as well as a total score, regardless of the order the numbers were repeated in. In this condition, Neijenjuis et al. showed a right ear advantage in scores with the mean difference in scores at 10% ($p<0.001$). This ear difference was not shown with the scoring procedure used in this study, likely because ears were not scored separately in the same manner.

Results are shown in Table 4 for the total scores calculated in the free recall condition. The younger subjects had mean scores of 70.2 for form A and 78.5 for form B. As expected, the fifth graders again showed higher mean scores at 79.2 for form A and 89.8 for form B in the DD task as compared to the third graders. There is an improvement of means scores in both groups from form A to form B ranging from about 9-11 items. As a group, the third and fifth grader mean scores are between 58-75% for DD. This percent score is very similar to the 60-72% range of total mean scores found by Neijenhuis et al. (2002) for the group of children from 9-12 years. They found the mean score to be about 65% for this age group, which would be similar to the 9-11 year-olds in this study.

Table 4. Means, standard deviations, and fmax ratio for binaural tasks **indicates significance at p<0.05

Grade	Test	Ear	Form A		Form B	fmax ratio
3rd (8-9 years)	CS	R	¯	12.4	11.2	1.60
			SD	4.8	3.8	
	CS	L	¯	11.4	11.1	1.13
			SD	4.7	5.0	
n=25	CS	Both	¯	23.9	22.3	1.34
			SD	8.9	7.7	
	DD	Both	¯	70.2	78.5	1.46
			SD	13.8	16.7	
5th (10-11 years)	CS	R	¯	13.7	13.3	1.72
			SD	3.8	2.9	
	CS	L	¯	14.6	13.8	1.10
			SD	4.0	4.2	
	CS	Both	_	28.3	27.09	1.19

n=23			SD	7.3	6.7	
	DD	Both	_ SD	79.2 16.3	89.83 13.5	1.45

The temporal tasks included PP and DP. The total score possible for each task was 20 items correct. Table 5 shows that the mean scores for the third graders were lower than for the fifth graders. This again shows the improvement by age that would be expected. The scores for third and fifth graders range from 35-66%, so the ceiling effect is clearly not occurring in this sample. This was a concern for PP on the previous MAPA, especially with the fifth graders.

The lowering of scores is consistent with the fact that both the PP and DP tasks now involve a series of four tones rather than three. In addition, reversals were not scored as correct, as was the protocol for the previous MAPA. This was also done in order to avoid a ceiling effect. However, when examining the full data set, including the data in a companion study to establish tentative norms, a score of 2 SD below the mean was not possible in PP since this was a negative number. Thus it appears that not scoring reversals as correct might be too stringent for this young age group (8-11 years). On the scoring protocol, many of the reversals were marked on the forms but not originally scored as correct. The PP portion was re-scored with reversals as correct. The outcome of this change in scoring and the tentative norms for the PP task are discussed in a companion study. For now it appears that reversals will be scored as correct, at least for the younger subjects, although this change is not reflected in the numbers reported here.

Table 5. Means, standard deviations, and fmax ratio for temporal tasks **indicates significance at p<0.05

Grade	Test	Form A		Form B	fmax ratio
3rd (8-9 years) n=25	PP	‾ SD	11.4 6.3	11.8 6.2	1.03
	DP	‾ SD	7.3 4.8	9.0 5.3	1.22
5th (10-11 years) n=23	PP	‾ SD	12.5 5.4	13.2 5.5	1.04
	DP	‾ SD	11.6 4.7	12.4 4.5	1.09

T-test examination of means

The paired t-test is used to compare the means of scores for related samples such as the tests on form A and B when given to the same individuals. It is used to determine if the differences in means between forms are significant. The use of a t-test indicated that there was not a significant difference on eight of the tests. However, the mSAAT combined test showed a significant difference at the .05 level (p=.043) and three tests, including the SINCA left ear (p=.008), DD (p=.000), DP (p=.002), were found to be significantly different at the .01 level. The results of the paired t-test are shown in Table 6.

Table 6. T-test results for tests on form A and B. **indicates significance at .05 level *indicates significance at .01 level N=48; df=47

A/B form test	Significance
mSAAT-right	.190
mSAAT-left	.118
mSAAT-ears combined	.043**
SINCA-right	.197
SINCA-left	.008*
SINCA-ears combined	.439
CS-right	.051
CS-left	.279
CS-ears combined	.071
DD	.000*
PP	.217
DP	.002*

For the combined mSAAT score, a .05 level of significance in the means for forms A and B is shown by a difference of about two words on a 50-word task, not a large practical difference. The .01 level of significance on three tests (SINCA left, DD, DP) might be taken to indicate a need for greater equivalence on these tests. The differences, however, in terms of practical testing considerations do not represent very large actual score discrepancies. The left ear SINCA results between form A and B show a mean difference of less than one word out of 20 given. A practice effect does not account for this because the form B results are poorer. So it would seem to be a matter of harder items. However, even though the practical difference is small, the items on both forms of the SINCA will be under revision in future research to find tests that show better correlations and less mean differences as discussed in the correlation section.

It appears in the other two cases (DP and DD) that a learning effect could explain the difference since the order of A and B were not randomized and form A was always given first. In these two tests, there are either limited or no practice items recorded on the CD for forms A and B. Thus, it seems likely that the differences indicate a learning effect with better scores on the second test (form B).

The mean difference between A and B forms for DP involves only about two items on a 20-item task. It is almost impossible to explain this difference on the basis of items since 17 of the 20 items on both tests are absolutely identical except for the randomized order. Among the remaining three different items of form B (LSSS, LSSL, and LSLS), one actually matches a similar pattern of form A (LSSS on form B matches LLLS on form A). Further, the remaining items do not appear that different. More practice prior to starting the DP task may be indicated since currently there is only one practice item.

The percent difference on DD is similar to DP in that it involves a difference of 9-11 items on a task of 120 items. It seemed that the use of the first 12 items on both tests for practice items might have eliminated this score difference while only reducing the scored items to 96. However, when this was calculated, the correlation between tests remained the same as well as the t-test significance of $p=0.000$. When the difference between means was divided by the total number of items, the percent difference only changed from 7.5% to 7.3% with the recalculation. Thus, not counting the practice items did not improve the mean difference between forms. Since both forms use the same eight digits, it does not seem possible that form A could be harder than form B. It seems more likely that learning occurs and that most people will perform better in the second testing session.

However, when considering both DD and DP, it is noteworthy that even though there are mean differences, this may not be a serious concern since the strong correlations between A and B forms for DD (r=.69) and especially DP (r=.85), as shown in Table 7, nevertheless indicate that the forms are highly related. Modest adjustments in practice items for DP could be considered.

Correlations between forms

Correlations between the A and B forms for the individual tests are shown in Table 7. When looking at mSAAT and SINCA, the correlations ranged from .17 to .28 and are disappointingly low as well as not statistically significant. However, the correlation for combined ear totals for each of the monaural tasks, mSAAT and SINCA, showed that the two forms are related more strongly when combined (.37 to .40). These correlations are statistically significant, but still only moderate in magnitude. For DD, CS, DP and PP, the correlations ranged from .69 to .89 and are very acceptable as well as statistically significant. The correlations show that forms A and B of all tests are significantly related, either when looked at by ear or, for mSAAT and SINCA, when ears are combined.

Table 7. Pearson correlation between test forms arranged by strength of correlation. N=48 **indicates significant correlation at .05 level

A/B form test	Correlation
PP	.89**
DP	.85**
CS-ears combined	.78**
CS-left	.74**
CS-right	.71**
DD	.69**
mSAAT-ears combined	.40**
SINCA-ears combined	.37**
SINCA-right	.28
mSAAT-left	.26
SINCA-left	.26
mSAAT-right	.17

The disappointing results on the SINCA led to item by item analysis on the form A test/re-test, which was examined in a companion study. The most reliable half of the SINCA items for each ear was given a double weighting for scoring. These changes moved test-retest correlations from .2-.3 to .5-.53. Thus, some improvement in this test may be possible, which would alter the mean scores as well. Eventually, the two reliable halves of the right and left ear SINCA items will be combined into one task for one ear on form A. Similar work will be done with the words on form B to provide a reliable task for the other ear on form A. When this is completed, more testing will be done to see if the changes improve the correlations. If it is successful, a similar procedure will take place to provide the SINCA items for form B.

Monaural, binaural, and temporal tasks

As has been discussed, the final MAPA will contain tasks in the monaural, binaural, and temporal areas. It is of interest to look at the means, standard deviations, correlations, and t-test results when these tasks are combined. The monaural tasks are the mSAAT and the experimental SINCA. The binaural tasks are CS and DD. The temporal tasks that were combined here are PP and the experimental Quick Tap test. The Quick Tap, an experimental task with only one version, is described and examined in a companion study. As mentioned previously, the Quick Tap, while given to all 48 subjects during form A, was only given to 25 subjects during form B. However, the results of the companion study suggest that the Quick Tap correlates well with the PP task so it is included in this portion of the study.

Table 8 shows the means, standard deviations, and t-test significance for the combined monaural, binaural, and temporal task scores. It also shows a composite score of all the tasks. The monaural tasks show a mean difference of 1.7 between forms (p=0.052). The low significance is not surprising considering the issues that have already been discussed concerning the SINCA. Again, when the practical difference is considered, nearly 2 items out of a combined total possible score involving 70 items does not seem to be a matter of large concern (2% of total score). The binaural tasks present with a mean difference of 8 items between the two forms, which is significant at the .01 level (p=0.000). This is also not surprising considering the apparent learning that seems to occur when the DD task is taken a second time. This difference represents 5% of the total score possible. The temporal tasks only show a .4 mean difference between forms (p=0.727). This result using the Quick Tap combined with the PP is very encouraging. It is also encouraging to see a mean difference in the composite score of only 3.7 (p=0.340). Thus, when the test is looked at as a whole, there is not a significant difference in the scores between form A and form B. In addition, the fmax ratio shows that there is not a significant difference in the standard deviations between forms.

Table 8. Means, standard deviations, t-test significance, and fmax ratio for combined monaural, binaural, and temporal tasks *indicates significance at .01 level

Tests	Form A		Form B	Sig.	fmax ratio
monaural n=48	 SD	39.1 5.9	37.4 5.4	.052	1.19
binaural n=48	 SD	100.5 19.4	108.5 21.1	.000*	1.18
temporal n=25	 SD	34.2 11.4	34.6 11.6	.727	1.04
Composite Total n=25	 SD	168.5 27.0	172.2 31.1	.340	1.33

The correlations for these tasks are shown in Table 9. All of the combined tests show correlations that are statistically significant. The monaural correlation of .46 shows a moderate but acceptable level of correlation. The binaural and temporal tasks show very high correlations between .81-90. When the composite scores are examined, there is a high and very acceptable correlation of .79.

Table 9. Correlations for combined monaural, binaural, and temporal tasks **indicates significance at .05 level

A/B form test	Correlation
monaural	.46**

binaural	.81**
temporal	.90**
Composite Total	.79**

Question 2

This section compares the means and standard deviations of four tests, mSAAT, CS, DD, and PP on the current MAPA compared to previous MAPA versions. This wad done to determine if scores compare favorably to previous versions and to see if the ceiling effect was overcome, which was a concern in the previous MAPA. The tasks that were modified for the Beta III MAPA to make them more difficult include CS, DD, and PP. The mSAAT was modified between the first two Beta versions of the MAPA.

The scores are shown as percent means and standard deviations in Table 10. The present third grader scores are compared to the scores found by Domitz (1997) with the Beta I version of MAPA. The present fifth grader scores are compared to the scores of Shiffman (1999) with the Beta II version of MAPA. Part of the modification for the Beta III MAPA included presenting PP binaurally. Thus there is only one score for each grade to compare to the two tasks shown by the previous MAPA. In addition, the current scoring procedure used for DD did not show significant differences by ear, as discussed previously. Thus, the total percent score is compared to the previous right and left ear scores. The means and standard deviations for the present study, while not displayed in percent format, are also shown in Tables 3, 4, and 5.

Table 10. Percent means and standard deviations compared for Beta I, II, and III version of the MAPA.

Test		Third grade		Fifth grade	
		Beta I	present	Beta II	present
mSAAT-R	SD	65.9 10.26	47.4 9.64	78.33 10.98	49.6 13.25
mSAAT-L	SD	61.09 11.53	47.6 9.44	77.0 7.65	47.13 10.02
PP-R	SD	78.0 21.1	57.2 31.7	97.98 3.44	62.61 26.96
PP-L	SD	81.7 20.5		97.68 2.3	
DD-R	SD	92.3 7.8	58.5 11.49	95.63 5.55	65.98 13.57
DD-L	SD	78.8 15.5		90.63 8.80	
CS-R	SD	89.8 13.4	41.5 15.93	98.33 5.77	45.8 12.68
CS-L	SD	68.1 23.5	38.1 15.5	99.17 2.89	48.7 13.4

The mean percent scores for all tasks are lower in the present study compared to the other versions of the MAPA for both third and fifth graders. The lower scores in PP and DD can be explained by increased difficulty of the tests (four series tones for PP and digit triplets for DD). The lower scores in CS can be explained by the change in instructions which ask for both sentences to be repeated instead of one. The Beta III mSAAT is different from the Beta I version, which would explain some of the changes. The Beta II mSAAT is the same test. However, the sample size for the Beta II MAPA data was low and only contained 12 subjects.

The results from Table 10 indicate that the ceiling effect was clearly not occurring on any of the original tasks with the Beta III MAPA. The current scores, while lower than the other versions, are still considered reasonable for these age groups.

CHAPTER IV

Summary and Conclusions

The purpose of this study was to determine 1) if forms A and B of the Beta III MAPA were equivalent and 2) if the modifications made to some of the tasks were sufficient to overcome the ceiling effect and produce reasonable means and standard deviations for the Beta III version of the MAPA.

Means, standard deviations, Pearson correlations, and t-tests were used to examine the scores for the first question. First, the individual scores of mSAAT, SINCA, DD, CS, PP, and DP were examined. The scores as groups of monaural, binaural, and temporal tasks as well as a composite score were also examined for the entire test.

When looked at individually, the two forms for the individual ear monaural tasks mSAAT and SINCA do not correlate in a significant manner. However, when the ears are combined, the forms are more strongly related and statistically significant (.37 to .40). The results of the paired t-tests show that the mean differences on form A and B scores are not significant, either when looked at by ear or when combined. As combined monaural tasks, the msAAT and the SINCA show a correlation of .46, which is moderate in magnitude, but statistically significant. Thus, these forms are passably acceptable to use as equivalent forms. However, to improve the correlations, the SINCA task will be modified and studied through future research. This future research will determine final form equivalency in the monaural tasks.

The binaural tasks, CS and DD, show correlations between .69-.78 for forms A and B. This shows an acceptable relationship between forms for both tasks. The paired t-test showed significant mean differences between the scores on DD for form A and B. However, these differences seem to be attributable to learning that occurs on this task. As combined binaural tasks the correlation between forms is .81. This shows that on binaural tasks, the forms are highly related and exhibit good equivalency. The t-test for the combined binaural tasks shows a significant difference in means, but this is again probably due to the learning on DD as described.

The temporal tasks, PP and DP, show the highest correlations between forms at .89 and .85 respectively. This shows a strong relationship and good equivalency between forms. The DP does show a difference between mean scores that is significant at the .01 level ($p=0.002$). The reason for this is not well understood since the majority of the items between form A and B are the same and only presented in a different randomized order. The percent mean scores range from about 57-66% on the PP task and 37-62% on the DP tasks. Thus the ceiling effect is clearly not occurring in this sample. For the PP task, it was necessary to score reversals as correct for the younger subjects in the sample (8-9 year-olds) in order to establish tentative norms. Reversed items on the DP tasks were not scored as correct.

A combined temporal task score was found using the PP and the experimental task, Quick Tap, as suggested by a companion study. The correlation between forms was .90, which shows a high relationship and good equivalency. In addition, the difference between mean scores was not significant ($p=0.727$).

These findings show that correlation between forms is best on binaural and temporal tasks with some mean differences on forms A and B scores which are probably attributable to a learning effect. Therefore, reasonably good form equivalency was found. The monaural tasks show the lowest level of significant correlations between forms. However, work is being done on at least one monaural task to improve future correlations. For now, all tests on both forms demonstrate at least some measures of equivalency. Future research will determine if changes to monaural tasks will improve correlations and form equivalencies.

APPENDIX A

Literature Review

In recent years, many have promoted the necessity of having good screening tools and diagnostic measures for APD. In order to achieve this, the instruments used to screen and diagnose APD should have good validity and reliability. The work Schow and Domitz (2000) did on the original MAPA, with factor analysis results on four behavioral tests, showed that the MAPA yields a reasonable three or four factor structure for measuring APDs. With the two forms of the Beta III MAPA, it is important to re-establish the validity and reliability. Cacace and McFarland (1995) encouraged researchers to give the necessary attention to test reliability and to change factors that have resulted in poor reliability in past APD tests. This chapter will include a discussion of test equivalency, and factor analysis. It will also give a description of the MAPA tests as well as the changes made to the Beta III MAPA. Finally, it will discuss other tests and protocols currently being used or researched.

Test equivalency

Equivalent or alternate forms of tests are used for two main purposes. One purpose is to show test reliability. Alternate test forms show equivalence and reliability when the two forms show similar mean performance by the group of subjects and high correlation between the two forms. Anastasi (1982) states that “The correlation between the scores obtained on the two forms represents the reliability coefficient of the test” (p. 111). Anastasi (1982) also urges that the two forms are indeed parallel in terms of difficulty, format, time, etc. When the forms are found to be equivalent and highly correlated, then alternate forms are useful for many test purposes.

Many testing situations require alternate forms. For example, if a test needs to be re-administered after a short time interval, information on specific test items will not be available. Alternate forms can also be used in follow-up studies or to measure subject learning. In the case of the

Beta III MAPA, the two forms can be useful to measure the progress in the treatment of APDs.

"Ideally, alternate forms of a test are interchangeable in use" (American Psychological Association, 1985, p. 31). In this situation, it would not matter whether form A or B is used. In its final form, the alternate forms of the MAPA should be able to be used interchangeably.

Factor analysis

Factor analysis is another tool that can be used to help define precise areas within APD. According to Schow and Chermak (1999), "In addition to testing models of central auditory processing and CAPD to establish their construct validity, factor analysis of central auditory performance scores provides an important method by which we can group the underlying deficits that purportedly comprise CAPD" (p. 141).

In an attempt to provide a reliable screening tool with factor analysis results, Domitz and Schow (2000), administered a battery of four behavioral APD tests to school-aged children. The test battery was selected following Musiek & Chermak's 1994 suggestions. The test battery was named the Multiple Auditory Processing

Assessment (MAPA) and included the following: monaural Selective Auditory Attention Test (mSAAT), Pitch Patterns (PP), Dichotic Digits (DD), and Competing Sentences (CS). Factor analysis on the four tests showed that they loaded into four distinct categories, which Domitz and Schow named. The mSAAT scores loaded into one factor which they called monaural separation/closure (MSC). The PP scores loaded into another factor which was called auditory pattern/temporal ordering (APTO). The loading factor for the DD scores involved binaural integration (BI). Finally, the CS scores constituted tasks requiring binaural separation (BS). Domitz and Schow determined that three of the most common and important ASHA 1996 test categories were represented in this test battery (binaural dichotic tasks, monaural tasks, and temporal tasks). Binaural interaction, speech recognition, and localization/lateralization were within the ASHA categories but not represented in the MAPA. Nevertheless, they felt that the MAPA was an appropriate test according to the 1996 ASHA guidelines and more work would be needed to demonstrate the need to test in the last three areas.

MAPA concerns

While much work has been done on the MAPA, McFarland

and Cacace (2002) stated concerns with the MAPA that they felt required further research. One area is to determine if the MAPA actually tests distinct aspects of CAP as suggested by the factor analysis. They mentioned the research done by Domitz and Schow (2000), on the factors shown in the SCAN as compared to the MAPA. Domitz and Schow (2002) reported that the SCAN showed two loading factors in contrast to the one factor reported by Amos and Humes (1998). Schow et al. (2002) agreed that further research needs to be done on the MAPA in the area of factor analysis. The purpose of the further research also includes gathering tasks for the MAPA battery that test as many areas as possible that are described in the 1996 ASHA document. Schow et al. mentioned the desire to have a battery with a number of tests for each of the underlying traits of auditory processing.

McFarland and Cacace (2002) also questioned the auditory-only tasks on the MAPA. They felt that this limited the ability to distinguish auditory-specific effects from more general aspects, such as cognition. They felt that if the tasks truly do measure auditory only aspects, these tasks should not be able to predict performance on other sensory tasks. In an earlier article Cacace and McFarland

(1998) stated that “the primary deficit with CAPD should be manifested in tasks requiring the processing of acoustic information, and should not be apparent when similar types of information are processed in other sensory modalities” (p. 356). Schow et al. pointed out that they were cautious to only use auditory tasks rather than involving the other senses, such as vision. However, they invited further research to be done in this area.

Test Description

The Beta III MAPA contains four tests from the original MAPA. The four tests are the mSAAT, a monaural task, DD and CS, binaural tasks, and PP, a temporal task. The final version of the MAPA will eventually contain at least two tasks in each of the monaural, binaural, and temporal areas.

The Selective Auditory Attention Task (SAAT) was a binaural test developed by Cherry in 1980. The test requires the individual to listen for the primary stimuli (words selected from the Word Intelligibility by Picture Identification or WIPI list) which are embedded in competing background noise. The earlier MAPA and the Beta III MAPA use a monaural version of this test with both the stimulus and the competing noise going to the same ear. Thus, in these test batteries, the test is referred to as mSAAT. The mSAAT contains 25 items per ear. The test time, including instructions, is approximately three minutes for each ear for a total of six minutes.

The Pitch Pattern test (PP) introduced by Pinhiero in 1977, randomly introduces high and low pitch qualities in a three-tone series which must be identified. Willeford and Burleigh (1985) reported that the test allowed multiple response modes. The subjects could verbalize, hum, sing, or manually point (high or low) to make their responses. The current protocol also allows this multiple response mode. This is not stated explicitly in the instructions recorded on the CD as modeling different response modes might only confuse some of the children. Instead, the instructions were kept simple and all kinds of responses were allowed. It was found that some of the children would naturally sing their answers. Later testing can explore humming or singing if necessary. The changes in the PP task for the Beta III MAPA will be discussed in the next section.

The Dichotic Digits test (DD) introduced by Musiek in 1983, presents four numbers simultaneously to the listener, two in each ear. The subject is required to repeat all four numbers aloud in a free recall manner. The order the numbers were repeated was not taken into account when scoring DD. The Beta III MAPA has presentation changes as well as a revised scoring procedure for DD. These changes will be discussed in the next section.

The Willeford Competing Sentences test (CS) presents two sentences, one to the right ear and one to the left ear, concurrently. The subjects only repeated the left or right ear sentence as directed while ignoring the opposing ear. There have been instruction changes in the Beta III MAPA for the CS task that will be discussed in the following section.

Changes to the MAPA

The revised version of the MAPA (Beta III MAPA) has been developed and recorded on compact disc (CD) by Auditec of St. Louis. The revised MAPA includes test information for a form A and form B version of the test. As mentioned, the current MAPA has the same tests as the earlier version. Some changes have been made to make the tasks more difficult in order to avoid the ceiling effect that was occurring with the MAPA and other APD tests. For example, Neijenhuis, Stollman, Snik, and Broek (2001) reported on a battery of tests that included pitch and duration pattern tests. They also found the ceiling effect was occurring on these tests, depending on the age of the subjects.

There are variations on three of the Beta III MAPA tasks. One variation is on the CS task. In his early work on competing sentences, Willeford (1978) mentioned the possibility of testing the patients’ ability to repeat both sentences. He said that in this case, both sentences should be presented at the same testing level of 50 dB HL. He reported that normal adult subjects should be able to do this task easily. The Beta III MAPA, following this advice, required the subjects to listen to the two sentences presented simultaneously in the left and right ears and repeat both sentences. They were directed to repeat either the right or the left ear first (Chermak, personal communication, March 5, 2003). In the Beta III MAPA, which included 15 sentences for each ear, the instruction and test time was approximately three minutes for each ear for a total of six minutes.

The dichotic digits task has been modified to present number triplets in each ear instead of two numbers. The subjects were directed to repeat items from the right ear first for ten items and from the left ear first for the other ten items as recommended by Moncrieff and Musiek (2002). They recommended the directed response mode because it would control attentional strategies and might provide information about laterality. This will be discussed in more detail in the section containing results from other test batteries. Scoring for DD was done in two ways in the current research. The first way was a total score for all correct numbers in a free recall situation. The other way provided a score for the right and left ear items repeated first as directed. Thus, the order that the subjects repeated the numbers was taken into consideration in the alternate scoring procedure. In the Beta III MAPA, this task took approximately four minutes.

The final task variation was with the PP task. The main difference was that four tones of high and low pitches intermixed were presented instead of three tones to make the test more difficult. In the original MAPA, the scoring for PP allowed exact reversals, such as high-low-high in the place of low-high-low to be scored as correct. The scoring of the PP task was initially changed to avoid the ceiling effect. In this change, reversals of the tones were not scored as correct in the Beta III MAPA. Later, when examining the results, it was found that the scores for the younger children were very low. The scoring was modified to count reversals as correct. This improved the scores, at least for the 8-9 year olds. Currently, reversals will be scored as correct, at least for the younger children. The final PP change is that there are now twenty total PP tasks given binaurally instead of the thirty items used previously for each ear. Statistical analysis was done on the data gathered by Domitz and Schow (1997) with regard to the number of pitch pattern items needed. The analysis showed that each item in the list was equivalent to any other item in terms of difficulty. The correlation between 15 and 30 tasks was .92. The correlation between 20 and 30 tasks was .96. Thus, the results showed that twenty items were sufficient to provide a reasonably valid score. The test and instruction time for this task on the Beta III MAPA was approximately four minutes.

Experimental tests

There were four experimental tests recorded on the CD of the Beta III MAPA that were examined in detail by a companion study. Two of these tests, the Speech in Noise for Children and Adults (SINCA) and duration patterns (DP) have a form A and B. Thus, their means, standard deviations, and correlations were examined in this study. A description of these tests follows.

The SINCA is a monaural task given to the right and left ears. There are 24 total words to be repeated, which includes the first four words that were eventually found to be better when not scored, but used as practice items. There is a four-speaker babble in the background that gets progressively louder by 4 dB after each set of four words. Thus, the practice words are at +20 dB compared to the four speaker babble. For the last four words at the end of the task, the babble and signal words are at the same level. Scoring was completed by taking the total number right out of 20 (excluding the practice items) and subtracting this number from 18 to get the signal-to-noise ratio (SNR).

A variation of this strategy was used in the QuickSIN, a speech-in-noise test developed by Etymotic Research. In the QuickSIN, the SNR loss is found, which compares the subjects’ performance in noise to individuals with normal hearing. People with normal hearing require a +2 dB signal-to-noise ratio to correctly identify 50% of the key words in the QuickSIN sentences (QuickSIN test manual). The SINCA was looking for a SNR score, thus the 2 dB was subtracted from the 20 possible items, resulting in the number 18 which was used to find the SNR. In the Beta III MAPA, it took approximately three-and-a-half minutes to test both ears.

The DP task is very similar to the PP task in format. Each item included a series of four tones that varied in their presentation lengths (short or long). There were twenty total items presented binaurally. The subjects were required to correctly identify the pattern. Reversals for DP were not scored as correct for any age group. For the Beta III MAPA, this task took approximately four-and-a-half minutes.

Other test batteries

Neijenhuis, Stollman, Snik, and Broek (2001) administered a battery of seven auditory tests to 28 adults with normal hearing. Their purpose was to find tests appropriate for adults that could also be modified for testing older children. The battery included words in noise, filtered speech, binaural fusion, sentences in noise, dichotic digits, frequency and duration patterns, and backward masking. The dichotic digits task used presented three numbers to each ear. The subjects were not directed to repeat a certain ear first, rather they were allowed free recall of the items. The scoring was done for each ear and for both ears. The frequency and duration patterns tasks used combinations of three tones or three durations. In addition to the behavioral tests, a 30-item questionnaire on everyday listening situations was administered. The subjects reported their responses on the frequency of occurrence of the behaviors on a 4-point scale.

The results showed that the ceiling effect was occurring on the frequency and duration pattern tests with their adult subjects. The median scores were 98% for frequency and 100% for duration. Scores at the 10th percentile showed 89% for frequency and 90% for duration. The dichotic digits task showed that scores were higher in the right ear than in the left ear. The median scores for this task were 83% in the right ear and 76% in the left ear, which is significantly different ($p<0.05$) (2001). This is the only test that showed differences between ears.

Neijenhuis, Snik, Priester, van Kordenoordt, and van der Broak (2002) administered a test battery originally intended for adults to a group of children and teenagers. Of interest were the results for 75 children (ages 9-12) on dichotic digits and pattern tests since this is a similar age to the subject in this study. The dichotic digits were presented as number triplets to each ear in a free recall condition. The pattern tests included both pitch and duration. However, in this study, series of three tones instead of four were used. The dichotic digits showed a mean percent score between 60-70% for this age group. When the task was scored looking at the individual ear performance, the right ear scores were significantly better than the left ear ($p<0.001$). The percent mean difference was 10% between ears. The mean percent score for the pitch or frequency pattern test was between 80-90%. For the duration patterns test, the mean percent score was slightly lower in the 70-80% range. The ceiling effect was present in this 9-12 age group, especially in the frequency pattern test.

Moncrieff and Musiek (2002) administered three dichotic listening tests to normal (control subjects) and dyslexic 11-year-old children based on reports “that children with dyslexia, a language disorder that leads to reading difficulties perform poorly on dichotic listening tasks” (p. 429). One of these dichotic tests included dichotic digits, using pairs. Both free recall and directed reporting conditions (right or left ear first) were used. “The directed response format was used in this study to explore the effects of attentional bias in the free recall condition” and also to be able to compare this task to the directed response format used for competing word tasks (Moncrieff & Musiek, 2002, p. 433).

The overall results showed that the control subjects performed better than the children with dyslexia. In the free recall condition, there were not significant differences in the scores between ears (for both groups of children) or between normal and dyslexic children. In the directed response mode, the normal children performed significantly better than the children with dyslexia; however, there were not significant ear differences. The children with dyslexia showed a significant ear difference as they performed better when directed to report the right ear first. Moncrieff and Musiek also reported that the DD task using double digits might be too easy for 11-year-old children as scores in both groups of children were at or near the maximum performance level. Among their suggestions was to use three or even four digit pairs. Another suggestions was to randomly present digit pairs of one, two, or three items so that the stimulus interval would be uncertain to the listener (2002).

Moncrieff and Musiek (2002) suggested that using a directed response mode to control for attentional strategies produced results that may be “more reflective of hemispheric lateralization for language” (p. 436). They also suggested that while this directed response mode would sacrifice information about attentional strategies used by the listener, it would provide “more reliable laterality indices” (p. 436). Laterality is one of the ASHA recommended areas for testing in order to diagnose children with APD. With this in mind, Moncrieff and Musiek state that “test conditions that will produce the most valid measure of both direction and degree of lateralization are essential” (p.436).

SCAN-C development

The SCAN, a test for auditory processing disorders, was developed by Robert Keith to test children ages 3-11. It was originally published in 1986. The SCAN-C is a revision of this test. For the SCAN-C, Keith modified his competing word task based on an item-by-item analysis, added a competing sentences task, and gathered more normative data for children ages 5-11 (Keith, 2000).

Keith described the SCAN-C and some changes in a 2000 article. The SCAN-C offers raw scores, which are converted to standard scores, percentile ranks, and confidence intervals. This standardization allows the scores to be compared to other standardized tests on the same population, be it language or intelligence tests. However, there were concerns with the original SCAN concerning test re-test issues (Amos & Humes, 1998) and test performance by location (Emerson et al., 1997). Additionally, Amos and Humes also questioned some of the scoring methods.

The SCAN-C consists of four subtests: filtered words (FW), auditory figure-ground (AFG), competing words (CW), and competing sentences (CS). The AFG task is somewhat similar to the SINCA task on the Beta III MAPA as words are presented in the presence of speaker babble. The CS task in the SCAN-C is similar to the original MAPA where two sentences are presented and the subject is asked to repeat either the left or the right ear sentence.

The SCAN-C was administered to 650 children between the ages of 5-11 years. Demographic information gathered included age, gender, race/ethnicity, geographic regions, and parent education level. Information was also collected from the school about children diagnosed or receiving special services for ADD, ADHD, LD, speech or language problems, developmental delay, dyslexia, behavior disorder/emotionally disturbed, or any other health impairment. In addition, gifted children were also recognized. To be included in the sample, children had good speech intelligibility with few articulation errors, spoke/understood English proficiently, and passed a hearing screening at 500, 1000, 2000, and 4000 Hz. Children with severe disabilities were excluded from the study.

Keith reported that “raw score means increased and the standard deviations decreased with increasing age as expected, reflecting maturation of the central auditory nervous system” (2000, p. 441). Standard scores were developed, giving equal weighting to each subtest (a previous concern). In addition, ear advantage scores were obtained. In general, a right ear advantage was shown, which reflects the “left hemisphere dominance for language” (2000, p. 442). Keith also provided a cumulative prevalence of ear advantage to determine if the ear advantage scores fall within a normal range. Keith reported the subtest test-retest reliabilities, with a mean testing interval of 6.5 days, ranging from .67 to .78 for the children 8-11 years of age. Keith also reported the correlation between subtests for the SCAN and SCAN-C, which are FW=0.55, AFG=0.31, and CW=0.72. To answer the question about test location, Keith used a matched sample of 27 children tested in an audiometric sound-proofed booth or in a quiet room. T t-test showed no significant differences between the mean scores for the subtests or composite standard scores.

Attention deficit hyperactivity disorder

There has been much debate about APD and its relationship to ADHD (attention deficit hyperactivity disorder). Chermack, Hall, and Musiek (1999) mentioned research from one side that felt that APD and ADHD might reflect a single developmental disorder. They mentioned research on the other side that reflected the possibility of co-occurrence between APD and ADHD. While ADHD is a medical diagnosis and APD is an audiological diagnosis, Chermack, Somers, and Seikel (1998), pointed out that they both share a number of common symptoms, including “attention and listening problems, maladaptive behavior, distractibility, instruction-following difficulty, and increased time required to complete tasks" (p. 78). They also pointed out that while ADHD was traditionally looked at as an attention disorder, it is now portrayed as "a deficit in motivation and rule-governed self-control rather than an attention deficit" (Chermack et al., 1998, p. 78).

Consequently, these researchers surveyed audiologists and pediatricians on behaviors that reflect APD and/or ADHD. Their results showed that while there are common symptoms, the ranking of these symptoms differ with each diagnosis. For example, inattentive and distracted behavior were the only behaviors present for the diagnosis of children with both APD and ADHD at the level of one standard deviation above the mean. However, with the diagnosis of ADHD, these behaviors ranked as number one and two compared to six and seven on the ranking for APD. The first five items on the list for identifying behaviors for APD included the following as reported by Chermak et al., (1998, p. 80).

11. Difficulty hearing in background noise
12. Difficulty following oral instructions
13. Poor listening skills
14. Academic difficulties
15. Poor auditory association skills

These researchers felt it was significant that behavioral scales, as judged by pediatricians and audiologists, should differentiate APD from ADHD in this manner. However, they were still unclear about the relationship between APD and the predominately inattentive type of ADHD (ADHD-PI), as the behaviors are most similar to each other.

In an expansion study to answer this question, another survey was sent to pediatricians and audiologists asking them to rank behaviors found in children with APD and ADHD-PI by Chermak, Tucker, and Seikel (2002). The results showed six common behaviors in both disorders, which were “academic difficulties, distraction, poor listening skills, asking for things to be repeated, auditory divided attention deficit, and difficulty hearing in background/ambient noise” (p. 335). They found nine behaviors that showed up only on one list (APD or ADHD-PI) that serve to distinguish or differentiate the disorders. Overall, they felt that the set of behaviors that most highly characterize either APD or ADHD-PI were reasonably exclusive as the four most characteristic behaviors (asks for things to be repeated and poor listening skills for APD and inattentive and academic difficulties for ADHD-PI) were not similarly highly ranked on both lists (Chermak et al., 2002). They further mentioned that pediatricians consider ADHD-PI to relate to cognitive problems while audiologists characterize APD as an auditory-specific deficit. They suggested further research and collaboration between professionals to ensure proper diagnosis.

Electroacoustical measures

The Bruton conference recommended a minimal behavioral test battery as well as the use of electro/physiological/acoustic tests and neuroimaging studies. As mentioned, most of the research concerning APD has focused on behavioral tests. In 2001, Jirsa summarized the information provided by three current electrophysiological measures. The middle latency response (MLR) provides information about the maturation of the auditory system. However, it may be difficult to observe in children under the age of ten and myogenic activity might also influence the results. The P300 has also shown sensitivity to APD. However, the responses are highly variable and depend upon subject participation and attention to auditory tasks. Finally Jirsa (2001) reported that while mismatched negativity (MMN) showed potential, “the waveform is difficult to identify and measure with accuracy” (p. 156). Thus it is not yet clinically useful.

Jirsa (2001) suggested that for electrophysiological measures to be used routinely, “An objective electrophysiologic measure is needed that is relatively unaffected by myogenic activity, does not require active subject participation, can be completed relatively quickly without inducing patient fatigue, and can be readily identified with minimal calculations” (p. 156). Jirsa reported on researched done with maximum length sequences-auditory brainstem responses (MLS-ABR). The MLS uses a higher stimulus rates/second than other measures currently in use by using a pulse sequence stimulus. The final response pattern is obtained through mathematical derivations. Jirsa examined children, ages 9-13, diagnosed and compared their responses to a control group. Jirsa reported that there were statistically significant differences between the wave V latencies between each group, with the clinical group being longer. He suggested that the nature of the test would lead to information about subjects’ temporal processing abilities. However, he cautions that results should be interpreted with caution as limited research has been done so far on the MLS-ABR.