An Abstract of the Dissertation of

Tsai-hsing Hsia for the degree of Doctor of Philosophy

in the Division of Special Education and Rehabilitation

to be taken March 1993

Title: EVALUATING THE PSYCHOMETRIC PROPERTIES OF THE ASSESSMENT, EVALUATION, AND PROGRAMMING SYSTEM FOR 3 TO 6 YEARS: AEPS TEST 3 TO 6 YEARS (AEPS TEST)

Approved: Diane Bricker
Diane Bricker

Designed as a programmatic assessment and monitoring tool, the Assessment, Evaluation, and Programming System for 3 to 6 Years: AEPS Test 3 to 6 Years (AEPS Test) provides detailed information on the behavioral repertoire of preschool children with special needs. This study explored selected psychometric properties of this criterion-referenced measure.

A total of 82 children with and without developmental delays participated in the study. Data collection procedures involved videotaping children while they were participating in a variety of assessment activities. Children's performance was then assessed by coders using the AEPS Test. Children's scores on the AEPS Test were used to

examine (a) the agreement between total and domain scores when children's performance was coded by two independent observers, (b) the relationship between total AEPS Test scores and scores on all domains, (c) the extent to which the AEPS Test scores were influenced by children's age, and (d) the extent to which the AEPS Test scores were influenced by children's degree of impairment.

Interrater reliability was obtained by correlating test scores obtained by independent coders. Percentage agreement was also calculated as an index of score consistency. Internal consistency was examined by correlating six domain scores and total test scores. Analysis of variance was used to investigate the construct validity of the AEPS Test.

Findings indicate that the AEPS Test has satisfactory interrater consistency at both domain and total test levels. Analysis of internal consistency shows a strong relationship between individual domain and total test. Relationships varied between the six domains. The test is sensitive to performance differences between five year olds and three year olds, five year olds and four year olds. However, no significant score difference was found between three year olds and four year olds. The analysis of the relationship between the AEPS Test and children's degree of impairment shows that the test is sensitive to performance differences between normally developing children and children with developmental delays in the sample. With the exception of

the Social-Communication domain, the test does not discriminate performance differences between children who were less impaired and children who have more serious delays.

10) and 2 items inconsistently (1 times 2), his or her raw score in this particular domain is 22, and the percentage score is 78.6%.

Five qualifying notes are used to indicate the situation under which the assessment information is obtained. The qualifying notes include assistance provided, behavior interfered, reported data, modification or adaptation of assessment directions or item criteria, and direct test.

Research on the AEPS Test

The AEPS Test encompasses six major skill areas of young children and was designed to be used for programming and evaluation purposes. Extensive research has been conducted on the psychometrics and utility of the AEPS test to examine its reliability, validity, and functionality.

Research on the AEPS Test for Birth to 3

The technical aspect of the AEPS Test for Birth to 3 first was examined during 1983 with 10 normally functioning children and 22 children with disabilities. Findings generated by this preliminary study showed that the AEPS Test (then the EPS-I), in general, was a reliable and valid assessment instrument. Also, the research showed that the tool was able to provide useful information for use in

planning intervention programs (Bailey & Bricker, 1986). This first formal investigation of the psychometric properties of the AEPS Test provided encouraging results and led to a revision and modification of the test.

Extensive data were collected from 1984 to 1987 to examine the reliability, validity, internal consistency, and utility of the revised AEPS Test, and the results were published by Bricker, Bailey, and Slentz in 1990. One hundred and twenty-two children, including both normally developing children and children with various degrees of impairment in four states, were assessed by two independent observers using the instrument.

Interobserver reliability, represented by Pearson product moment correlations, ranged from .705 for the Social domain to .958 for the Gross Motor domain. The total test had a reliability of .966. All correlations were significant at the .001 level.

Test-retest reliability of the AEPS Test was evaluated by assessing 58 children on two different occasions with an interval of two weeks. Temporal consistency of the test ranged from .771 for the Social domain to .951 for the Gross Motor domain. The total test had a reliability of .955.

Concurrent validity was obtained by comparing children's scores on the AEPS Test with the same children's scores on two standardized, norm-referenced measures. The

correlations between children's (N = 34) AEPS Test total scores and the Bayley Mental Age and Motor Age were .931 and .881, respectively. The correlations between children's (N = 121) AEPS total test scores and the Gesell Maturity Age scores were .5098. The correlations between the AEPS Test and the two standardized developmental measures were both significant.

In terms of the internal structure of the AEPS Test, data collected on 77 children with no disabilities showed that the test items were arranged according to level of difficulty. In addition, the analysis of the scores obtained from 155 children showed that significant relations existed between strands and domains, indicating consistency within the six domains.

Interventionists from 23 sites responded to questionnaires designed to investigate the utility of the AEPS Test. In general, the users found that the test items contained clear criteria and were appropriate for designing intervention programs for young children. Also, most of the respondents reported that the AEPS Test could be administered in a reasonable amount of time, and the items could be used as important goals for the children they served.

The usefulness of the AEPS Test was investigated by Notari and Bricker (1990). Specifically, the study compared

the quality of the long-range goals and short-term objectives written by interventionists who used the AEPS Test and the quality of the goals and objectives developed by interventionists using other assessment measures. Forty-eight interventionists serving infants and young children with special needs in Iowa, Vermont, Oregon, North Dakota, and British Columbia participated in the study. The study showed that the interventionists who used the AEPS Test as an assessment tool were able to generate goals and objectives that were more functional, generic, measurable, and easily integrated within daily activities.

Extensive research on the AEPS Test for Birth to 3 shows that the test is a reliable and valid assessment instrument that can assist interventionists in developing quality goals and objectives for young children with special needs.

Research on the AEPS Test 3 to 6 Years

The AEPS Test for 3 to 6 Years was developed from 1984 to 1985, and it shares the same organizational structure as the AEPS Test for Birth to 3. The psychometric adequacy of the AEPS Test 3 to 6 Years (then the EPS-II) was investigated by Slentz in 1986. Based on the data from 53 children, the test was found to have high degree of interobserver consistency at the total test level (r = .94).

Interobserver reliability for the domains ranged from .60 for the Social domain to .94 for the Fine Motor domain. Interobserver consistency of the total test was high for a subgroup of children with disabilities (N = 18) at both the total test level ($\underline{r}$ = .96) and the domain level ($\underline{r}$ = .70-.96).

Test-retest reliability of the AEPS Test was examined by correlating two test results administered to 18 children within two weeks. High temporal stability was shown for the total test ($\underline{r}$ = .91) and three of the domains: Social-Communication ($\underline{r}$ = .77), Fine Motor ($\underline{r}$ = .86), and Cognitive ($\underline{r}$ = .91). Test-retest reliability was moderate for the Social domain ($\underline{r}$ = .50) and low for the Adaptive (then Self-Care) ($\underline{r}$ = .13) and Gross Motor ($\underline{r}$ = .07) domains.

Slentz (1986) further examined the internal structure of the AEPS Test (N = 53) and found a high degree of variability in the relationships between individual domains with low correlations between Social-Communication and Adaptive domains, and high correlations between Cognitive and Fine Motor domains. In addition, the results showed that all domains correlated strongly with the total test score with the exception of the Adaptive domain.

Construct validity (N = 53) was also included in Slentz' study. She found that the AEPS Test was sensitive to performance differences between three and four-year-olds,

non-delayed and at-risk children, and at-risk and children with mild impairments. However, the test was not sensitive to the performance differences between four and five-year-olds, and children with mild and moderate impairment. Slentz contributed the failure of the test to detect score difference between four and five-year-olds to the inclusion of young five-year-olds in the sample. No difference between the scores obtained by children with mild and moderate impairments may have been due to inconsistent assignment of children to various degrees of impairment categories.

Concurrent validity was the last psychometric category included in Slentz' study. Eighteen children's scores on the McCarthy Scales of Children's Abilities and the Uniform Performance Assessment System (UPAS) were correlated with their scores on the AEPS Test. Results showed a high correlation between the McCarthy General Cognitive Index and the AEPS Cognitive domain scores ($\underline{r}$ = .66). Also, the scores on the McCarthy Verbal scale and the AEPS Social-Communication domain scores were highly correlated at .72. A weak correlation existed between the McCarthy Motor scale and the AEPS Fine Motor domain ($\underline{r}$ = .35), and no relationship between the McCarthy Motor scale and the AEPS Gross Motor domain($\underline{r}$ = .06). The low consistency found in the motor scores was probably because the skills assessed by

the two measures were of different nature; one belonged to a more functional and pre-academic type of abilities, the other more related to diagnosing impairments in motor skills.

Strong congruent relationships were observed between the UPAS Communication and the AEPS Social-Communication scores ($\underline{r}$ = .87) and the UPAS Preacademic and AEPS Cognitive scores ($\underline{r}$ = .77). While the UPAS Social/Self-Help scores correlated highly with the AEPS Social domain scores at .75, a low relationship was obtained between the UPAS Social-Self-Help and the AEPS Adaptive scores ($\underline{r}$ = .30). The UPAS Motor domain correlated moderately with the AEPS Fine Motor domain ($\underline{r}$ = .52), but extremely low with the AEPS Gross Motor domain ($\underline{r}$ = .06). The different age groups for the which the tests were designed may account for the low correlations observed in the self-care and motor skills of the two tests. While the AEPS is designed for children developmentally from 3 to 6 years, the UPAS was developed to cover a greater life span of children, from birth to 6 years of age.

Slentz' (1986) study led to the revisions of the AEPS in 1988. In the Cognitive domain, wording in Strand B, "Demonstrates understanding of concepts " was changed to increase clarity, and one item was added to Strand E, "Recalling Events." In Strand A, Social-Communication

domain, two new items were added and one was eliminated. One item in the Social domain was removed.

The revised AEPS Test for 3 to 6 Years contains 241 test items in six developmental domains: Fine Motor, Gross Motor, Adaptive, Cognitive, Social-Communication, and Social. Research was needed to investigate the technical quality of this instrument. Reliability studies focused on the agreement between the test scores obtained by independent observers. Analysis of the internal structure of the test examined the relationships between total test and individual domains. Validity studies focused on the relationship between test scores and children's age and degree of impairment.

Research Questions

Following questions were posed to evaluate the psychometric properties of the AEPS Test for 3 to 6 Years:

1. What is the agreement between two raters independently using the AEPS Test to assess the same children?

2. Are the AEPS Test items arranged into domains that measure behaviors of different construct?

3. What is the relationship between children's age and performance on the AEPS Test?

4. What is the relationship between children's degree of impairment and performance on the AEPS Test?

CHAPTER III

METHOD

The AEPS Test was designed to be used by direct service personnel to collect information on the skill level of children, developmentally from 3 to 6 years, in six developmental areas. The purpose of this study was to evaluate the adequacy of the AEPS Test by examining selected psychometric parameters of the instrument.

Subjects and Setting

Two strategies were used to recruit subjects for the study. First, fliers were made and posted in the Eugene and Springfield community inviting young children aged three to six years to participate in a short-term afternoon play session in the Early Intervention preschool classrooms. The classrooms were located in the Center on Human Development at the University of Oregon. Parents who were interested in having their child participate in the play sessions were asked to contact the researcher or her assistants.

After children were recruited, a schedule was arranged so that two to three children attended a daily session.

Each child came two hours a day, two days a week, for two weeks. The purpose of the research was explained to the parents, and parental consent was obtained on the first day the child attended the class. Appendix B contains a copy of Informed Consent Form that parents were asked to read and sign. All of the subjects recruited using the first strategy were children without disabilities.

The second strategy involved visiting a number of community programs serving preschool children with and without special needs. The purpose of the study was explained to the classroom teacher, and informed consent was obtained from the parents through the teacher. The subjects recruited through this strategy included children without disabilities, children who were at risk for the development of a disability because of environmental reasons, and children with disabilities. Children without disabilities were recruited from the University of Oregon Child Care at the Erb Memorial Union, Westmoreland, Brown House, Congregational Church Preschool, Children's Playhouse, and Educational Environment, Incorporated. Lane County Relief Nursery was the setting from which the subjects who were at risk were obtained. Three early intervention programs in Eugene, Hillsboro, and Corvallis provided the subjects who had disabilities.

An identification (ID) number was assigned to each

child in the study. The first digit of an ID number represented the setting in which the subject was observed, and the rest of the digits indicated the order in which the child was assessed. For example, number 1001 was the first child who participated in the afternoon sessions held at the Center on Human Development; number 4003 the third child assessed at the Congregational Church preschool program.

A total of 82 subjects participated in this study. The subjects' mean chronological age was 55.40 months with a range from 39 to 75 months. Table 1 shows the number of subjects in each age category.

TABLE 1. Number of Subjects in Each Age Category

	Age				
	3	4	5	6	Total
Number of Children	23	30	24	5	82

All the 82 subjects participated in activities that assessed fine motor, cognitive, social-communication, and social skills. Scheduling difficulties and lack of equipment precluded all children from participating in gross motor and adaptive activities. Specifically, 63 subjects participated in activities that assessed gross motor skills,

and 59 children were involved in activities that targeted adaptive abilities.

Measures

In order to answer questions regarding whether the AEPS Test was sensitive to children functioning at different levels, teachers in the settings that served children who are at risk or have disabilities were asked to use the Degree of Impairment Definitions to rate the subject's level of impairment. In the Degree of Impairment Definitions, four levels of impairment, at risk, mild, moderate, and severe, are defined based on children's functional skills, adaptive behavior and amount of assistance required. The Degree of Impairment Definitons was used in the study of psychometric properties of the AEPS Test for Birth to 3 and was found to be reliable (Bricker, Bailey, & Slentz, 1990). Appendix C contains a copy of Degree of Impairment Definitions form.

A Demographic Information Form was used to record the subject's date of birth, gender, degree of impairment, assessment dates, and names of the activities in which the child participated. Appendix D contains a sample Demographic Information Form. The number of subjects in each age category is reported in Table 2, while the subject distribution by age and degree of impairment is reported in Table 3.

TABLE 2. Number of Subjects in Each Impairment Category

	Degree of Impairment					
	Non-disabled	At-risk	Mild	Moderate	Severe	Total
Number of Children	51	11	8	9	3	82

TABLE 3. Subject Distribution by Age and Degree of Impairment

Degree of Impairment	Age			
	3	4	5	6
Non-disabled	13	18	16	4
At-risk	4	4	3	0
Mild	3	4	1	0
Moderate	2	3	3	1
Severe	1	1	1	0

Note: N = 82

The AEPS Test Administration Manual and Data Recording Forms were used when observers collected data for the study. Copies of the AEPS Test Data Recording Forms can be found in Appendix E.

Procedures

This study used videotaping procedures to record children's performance on the AEPS Test. No research to date was available comparing the accuracy of assessment of children's abilities using videotape and assessments using live observation procedures. It may be that videotaping procedures had the advantage of allowing the coder to determine the existence of a skill more accurately by replaying the tape, thus increasing the accuracy of the assessment outcome.

Assessment Activity Plan

A set of assessment activity plans was developed to facilitate test administration. The assessment activity plans listed the materials needed for each activity, and listed all the possible AEPS skills that could be assessed as children participated in the activity. Following a prewritten activity script, the person who directed the assessment activity could set up the environment in a way to help ensure children would have opportunities to demonstrate targeted skills. The activities used in the assessment process were common in preschool settings and were of interest to most youngsters. Also, the activity plans were designed to permit assessment of skills across domains. For

example, in the snack activity, in addition to assessing adaptive skills such as eating a variety of food types and displaying appropriate dining skills, the children were asked to take five crackers from the container in order to observe counting skills, a Cognitive item. Also, engaging children in conversation provided opportunities for them to demonstrate skills in the Social-Communication domain, such as using verbs, adjectives, and asking questions.

The assessment activity plans included both routine activities (e.g., free play, snack) and planned activities (e.g., post office, washing babies). Appendix F contains a sample assessment activity plan.

Activity Schedule

The short-term afternoon sessions were attended by two to three children at a time. While children were attending the sessions, one interventionist directed the assessment activity and the other operated the videocamera. Because each child came two hours a day, two days a week for two weeks, eight hours of videotapes were collected for each subject.

A variety of activities were scheduled each day to motivate the children and provide opportunities for subjects to perform skills in all of the six developmental domains. For instance, story time was used to assess children's

preacademic skills, and snack time offered opportunities to assess eating skills. Playdough and water activities interested most of the preschoolers and provided opportunities to test communication and social skills. Appendix G contains a sample afternoon session schedule and a list of sample activities.

A different activity schedule was followed while collecting data in the community preschool programs. Two research assistants worked with two to three children at a time; one directed the activity while the other operated the videocamera. Each subject in the community program was videotaped three times in a one- or two-week period. Each day children were videotaped while participating in three planned and routine activities directed by the research assistants. Appendix H contains a community site sample data collection schedule.

Three information recording forms were completed by the research assistants. The Demographic Information Form was used to record subject's ID number, date of birth, gender, degree of impairment, videotaping dates, and the activities in which the child participated. A degree of impairment rating was obtained from the classroom teacher based on the Degree of Impairment Definitions. For each activity, a Videotape Information Form was used to identify the children on the videotape. The Videotape Information Form included

the site in which the activity took place, videotape number, name of the activity, videotaping date, subjects' ID numbers, and subjects' description (e.g., ID number 8001 is the girl with ponytail wearing a red and white striped T shirt). Subjects' names were not recorded on the form to ensure confidentiality. Appendix I contains a copy of Videotape Information Form.

Scoring Procedures

The standard AEPS scoring procedures were modified to meet the constraints of this study. Since the opportunity to observe the children's performance was limited, whether the behavior was functional and generalized (i.e., occurred in a variety of settings with different people and different materials) could not be addressed. Therefore, children's responses to items were evaluated only in terms of their presence or absence; inconsistency (i.e., I) was not used as a scoring option in the study.

Two symbols were used to represent children's performance on the AEPS Test. First, a plus ("+") was used when the child was observed to perform the skill independently. Behaviors only had to occur once to qualify for a plus. Second, a minus ("-") indicated the child was not observed to perform the skill or did not meet the criterion. For example, if the child did not use a tissue

to wipe his or her nose when a tissue was provided by the activity facilitator, the item was scored as a minus. Since the purpose of the study was to evaluate the reliability of the AEPS Test rather than determining the skill level of the subjects, no attempt was made to distinguish whether the child did not have the capability or opportunity to perform the skill or just simply refused to do what was asked. A minus was used for all the skills not observed by the coders.

All coders received written instructions regarding general and domain-specific scoring procedures. The objectives on the AEPS Test are arranged according to level of difficulty; children may not be observed demonstrating a lower level behavior once they master a more difficult skill. For instance, if a child is able to walk up and down stairs alternating feet without holding a handrail or wall, this child may not be seen walking up and down stairs without alternating feet and with support. In order to measure children's highest ability, if a child was able to independently perform a more advanced skill, the lower level skills that lead to the more advanced skill did not need to be observed to be scored as a plus. All items appropriate for this scoring method were listed in the scoring procedures. Appendix J contains a copy of scoring procedures.

Training of Coders

Seven students in the Early Intervention/Communication Disorders Sciences graduate program served as the coders for the study. The coder's responsibility was to view the videotapes individually and score the children's performance using the modified AEPS scoring procedures.

Only the information on the Videotape Information Form (e.g., testing site, tape number, activity, testing date, children's ID numbers and description) was available to the coders. The coders were blind to the subjects' degree of impairment and age. Some of the coders served as assessment activity facilitators during the videotaping process; however, assignment of coders was done in such a way that a coder never scored a tape on which she had served as the activity facilitator.

All coders received a packet of reading materials including an AEPS Test test manual, a set of assessment activity plans, AEPS Test Data Recording Forms, and scoring procedures. A 10-hour coder training session was required to reach a satisfactory level of agreement (i.e., 80%) before the coders began to score the participating children's performance. The training consisted of reviewing the test manual and items, scoring procedures, and practice sessions. Videotapes of six children without disabilities,

ages 3, 4, and 5, participating in the assessment activities were used as training tapes in the practice sessions. These six children were not included as subjects for the research.

Initially, coders were trained to focus on one domain at a time. The children's behavior tapped by a particular domain of the AEPS Test was pointed out by the researcher, and the coders and researcher reviewed the scoring criterion together. For example, the ability to bounce balls is one of the items in the Gross Motor domain, and the criterion requires the child to bounce a ball at least eight inches in diameter at least twice using the palm of one hand. Items and their criteria in each of the six domains were reviewed, and the coders were reminded to check the test manual if they were not sure whether the child's response met the item criterion.

Coders were required to take language samples in order to score the items in the Social-Communication domain. Guidelines for observing and recording the behavior in this domain (p. 247, AEPS Test Administration Manual) were modified slightly to permit scoring from videotapes. Rather than recording a child's utterances for 10 to 15 consecutive minutes, coders were required to record verbatim everything the child said during all the activities in which the child participated. Videotapes could be stopped and replayed to allow the coders sufficient time to record children's

utterances. Appendix K contains guidelines for scoring the Social-Communication domain.

Interrater agreement at the domain level was calculated at the end of each training session. An index of agreement was obtained by dividing the number of agreements in a domain by the number of agreements and disagreements (i.e., total number of items in a domain). After an 80% or higher interrater reliability was achieved in all six domains, the coders were then trained to score a child's performance across domains simultaneously. For example, while taking a child's language samples, his or her cognitive skills regarding spatial and color concepts could be evaluated at the same time. Discrepancies in scoring were discussed until a consensus was reached.

Training was concluded after an 80% or higher agreement in the six domains was obtained. Coders were assigned randomly in pairs to assess the same child. However, the videotapes were viewed and coded independently. Each child's performance was scored by two coders.

CHAPTER IV

RESULTS

A total of 82 subjects in 10 sites participated in this study which examined selected psychometric properties of the AEPS Test. A child's performance on the AEPS Test can be represented by domain scores and percentage scores. A total test score can be obtained by summing all of the six domain scores. In this study, domain raw scores and total test scores were used in answering the research questions.

This chapter presents the study's results. Specifically, the interrater reliability, internal structure, and construct validity of the instrument were examined.

Interrater Reliability

The first research question was: What is the agreement between two raters independently using the AEPS Test to assess the same behaviors of the same children? One method of reducing error variance of a test is to use standardized procedures for administration. Children being tested interact with the same materials, are given the same

instructions, and are scored using the same criteria (Bailey & Brochin, 1989). Using standardized procedures allows the scores obtained from different children to be compared. The AEPS is not a standardized test in that it is administrated by observing young children's functional skills in a familiar setting. Test examiners are free to use the materials that match children's interest and age level. Adaptation of the testing procedures also is permitted in order not to penalize the child for his or her disabilities. Test results, then, are not used to compare performance across children but to determine their level of functioning and to design appropriate intervention content and strategies.

Since the AEPS Test does not employ standardized test administration procedures, one way to increase the interrater reliability of the test is to operationalize the behavior being assessed and specify the criterion for correct performances. A test is of little value if the items are worded ambiguously, making it difficult to determine whether the child's performance satisfies the criterion. In the process of developing the AEPS Test, an effort was made to ensure that the wording of the test items was objective and limited personal judgement was needed to interpret children's performance. Each of the test items and their criteria was written in an observable and

measurable manner so that children's skill could be scored reliably.

For this study, interrater reliability was examined by correlating children's domain and total test scores obtained by two independent scorers. Data for all 82 subjects, 62 children without disabilities and 20 children with disabilities, were included in the interrater reliability analysis. However, not all of the children had opportunities to participate in activities that tapped skills in the Gross Motor and Adaptive domains. Sixty-three children (50 with no disabilities and 13 with disabilities) were included in the Gross Motor domain analysis, and 59 children (41 with no disabilities and 18 with disabilities) were included in the Adaptive domain analysis.

Subjects' performance was video recorded, and the videotapes were viewed and scored by two coders independently using the AEPS Test criteria. Children's performance was scored by coders as either plus ("+") or minus ("-") based on the criteria stated in the AEPS Test manual and the modified research scoring procedures. Interrater reliability was calculated for domains and for the total test by calculating the percent of agreement between independent coders. The number of agreements between two coders was divided by the total number of test items to obtain percentage agreement. In addition,

interrater reliability also was evaluated by correlating the total test scores and individual domain scores obtained by two coders.

Table 4 presents the average percent agreement for domains and for total test for total group, subjects with disabilities, and subjects without disabilities.

TABLE 4. Average Percentage Agreement for the AEPS Test Domain and Total Test Scores Obtained by Independent Coders: 95% Confidence Intervals in Brackets

	Group		
Domain	Total	Non-disabled	Disabled
Fine Motor	.92 (N = 82) [.90 - .94]	.91 (N = 62) [.88 - .94]	.95 (N = 20) [.91 - .99]
Gross Motor	.90 (N = 63) [.88 - .92]	.91 (N = 50) [.89 - .93]	.92 (N = 13) [.89 - .95]
Adaptive	.94 (N = 59) [.93 - .95]	.94 (N = 41) [.92 - .96]	.94 (N = 18) [.92 - .97]
Cognitive	.90 (N = 82) [.88 - .92]	.88 (N = 62) [.86 - .90]	.94 (N = 20) [.91 - .96]
Soc-Comm.	.90 (N = 82) [.88 - .92]	.91 (N = 62) [.89 - .93]	.87 (N = 20) [.83 - .90]
Social	.87 (N = 82) [.85 - .89]	.86 (N = 62) [.84 - .88]	.90 (N = 20) [.87 - .93]
Total	.90 (N = 82) [.89 - .91]	.90 (N = 62) [.89 - .91]	.91 (N = 20) [.89 - .93]

As shown in Table 4, the percent agreements between two independent coders were very high for the total test for the total group and for the two subgroups; an agreement of .90

was obtained for the total group, and .90 and .91 for the non-disabled subjects and disabled subjects, respectively. Similar results were obtained at the domain level. Very high agreements were found for all the subjects (.87 - .94), subjects with no disabilities (.86 - .94), and subjects with disabilities (.87 - .95). The highest agreement (.95) was found in the Fine Motor domain for the group with disabilities, and the lowest agreement was found in the Social domain for the children without disabilities (.86).

Pearson product-moment coefficients were used to correlate the domain and total test scores obtained by two coders. The significance of the correlations was tested, and the size of Pearson's r was interpreted based on the guidelines provided by MacEachron (1982): very low ($\underline{r}$ = .00 - .20), low ($\underline{r}$ = .20 - .40), moderate ($\underline{r}$ = .40 - .60), high ($\underline{r}$ = .60 - .80), and very high ($\underline{r}$ = .80 - 1.00). The results for the total group, children with disabilities, and children without disabilities are presented in Table 5.

A very high degree of consistency was demonstrated for the total group and two subgroups for the total test ($\underline{r}$ .93 - .97), as shown in Table 5. At the domain level, for the total group, very high correlations were found on five of the six domains: Gross Motor, Adaptive, Cognitive, Social-Communication, and Social ($\underline{r}$ = .82 - .97). A correlation of .75 was obtained for the Fine Motor domain, the lowest for all the domains for the total group.

TABLE 5. Pearson Product Moment Correlations for the AEPS Test Domain and Total Test Scores: 95% Confidence Intervals in Brackets

	Group		
Domain	Total	Non-disabled	Disabled
Fine Motor	.75* (N = 82) [.64 - .83]	.74* (N = 62) [.60 . 84]	.63* (N = 20) [.26 - .84]
Gross Motor	.83* (N = 63) [.73 - .89]	.79* (N = 50) [.66 - .88]	.88* (N = 13) [.64 - .96]
Adaptive	.82* (N = 59) [.71 - .90]	.68* (N = 41) [.47 - .82]	.81* (N = 18) [.55 - .93]
Cognitive	.95* (N = 82) [.92 - .97]	.92* (N = 62) [.87 - .95]	.93* (N = 20) [.83 - .97]
Soc-Comm.	.96* (N = 82) [.94 - .97]	.84* (N = 62) [.75 - .90]	.95* (N = 20) [.88 - .98]
Social	.83* (N = 82) [.75 - .89]	.79* (N = 62) [.67 - .87]	.59* (N = 20) [.20 - .82]
Total	.97* (N = 82) [.95 - .98]	.93* (N = 62) [.89 - .96]	.97* (N = 20) [.92 - .99]

*p<.01, two-tailed test

The correlations between coders were lower for the non-disabled group, ranging from .68 - .93. For this subgroup, moderate consistency was obtained in the Adaptive domain ($\underline{r}$ = .68), and the correlations of three domains, Fine Motor, Gross Motor, and Social, fall in the high range ($\underline{r}$ = .74 - .79). Scores in the Cognitive and Social-Communication domains were highly correlated at .92 and .84, respectively.

For the group of children with disabilities, the

correlations were very high in four of the six domains: Gross Motor, Adaptive, Cognitive, and Social-Communication ($\underline{r}$ = .81 - .95). Similar to the non-disabled group, a lower agreement was obtained for the Fine Motor and Social domains for the children with disabilities; correlations in both domains are moderate.

Internal Structure

Internal structure of the AEPS Test was examined in order to answer the question: Are the AEPS Test items arranged into domains that measure different groups of behavior? The correlations among six domain scores and the total test scores were calculated to examine the relationship between each domain and total test and the degree of interdependence within each domain.

The test scores of 50 children who participated in the assessment activities that addressed all six AEPS domains were used to analyze the internal structure of the test. Table 6 presents the means and standard deviations for the six domains and total test scores for the total groups and two subgroups.

As shown in Table 6, the six domain scores and total test scores obtained by the children with no disabilities are consistently higher than those obtained by the children with disabilities. The greatest difference between the two groups exists in the two domains that contain the most

TABLE 6. Means and Standard Deviations for AEPS Domain and Total Test Scores for the Total Group and Two Subgroups

	Total (N=50)		Group: Non-disabled (N=38)		Group: Disabled (N=12)	
Domains/Total	M	SD	M	SD	M	SD
Fine Motor	4.38	2.45	5.00	2.37	2.42	1.31
Gross Motor	5.72	2.62	6.36	2.43	3.58	1.93
Adaptive	8.04	3.55	9.13	3.20	4.58	2.11
Cognitive	31.74	17.14	38.61	13.39	10.00	5.19
Soc-Comm.	35.64	11.07	40.37	5.72	20.67	10.69
Social	8.56	4.10	9.92	3.52	4.25	2.56
Total	94.08	36.00	109.42	24.10	45.50	20.80

Note: N = 50

items: Cognitive and Social-Communication.

Based on the scores of the 50 subjects, Kuder-Richardson Formula 20 (Kuder & Richardson,1937) was used to calculate the internal consistency of the AEPS Test. KR-20 is a procedure for calculating the extent to which one can generalize to other sets of similar test items. This procedure requires only one test administration, and it is appropriate for dichotomously scored test items (Salvia & Yesseldyke, 1991). According to Nunnally (1978), investigations of internal consistency involves averaging correlations among test domains, and the size of the reliability coefficient depends on both the internal consistency and the number of test items. The results indicate that the test has a KR-20 estimate of .98 (241 items) for the total test and the reliability for the six domains were: (a) Fine Motor (14 items), .75; (b) Gross Motor (18 items), .64; (c) Adaptive (39 items), .79; (d) Cognitive (88 items), .97; (e) Social-Communication (49 items), .96; and (f) Social (33 items), .74. Salvia and Yesseldyke suggest that, for tests designed to provide instructionally relevant information, the minimum standard of reliability should be .90. Findings suggest that total test and two of the domains, Cognitive and Social-Communication, have satisfactory internal consistency.

To examine the extent to which the six domains measure related constructs, Pearson product-moment correlations were

calculated between each domain and total test. Three Pearson product-moment matrixes for the total group and two subgroups are presented in Tables 7, 8, and 9. The size of the Pearson's r results was interpreted based on the guidelines developed by MacEachron (1982).

As shown in Table 7, for the total group, the correlations between individual domain and total test were significant and ranged from very high for the Cognitive domain (r = .95) and Social-Communication domain (r = .90) to moderate for the Gross Motor domain (r = .67). The correlations among the six domains ranged from high between the Cognitive and Social-Communication domains (r = .78) to low between the Adaptive and Gross Motor domains (r = .46).

Table 8 presents the intercorrelations between domains and total test for the non-disabled group. Similar to the data obtained for the total group, the total test was highly correlated with the Cognitive domain (r = .93) and Social-Communication domain (r = .75). The Adaptive and Gross Motor domains correlated moderately with the total test. For the non-disabled group, the correlations among the six domains ranged from moderate between the Social-Communication and Cognitive domains (r = .59) and Fine Motor and Adaptive domains (r = .60) to low between the Social-Communication and Adaptive domains (r = .22).

Table 9 presents the intercorrelations between domains and total test for the disabled group. Total test

TABLE 8. Intercorrelations Among Domain and Total Test Scores for the Non-Disabled Group: 95% Confidence Intervals in Brackets (N=38)

Domains/ Total	Fine Motor	Gross Motor	Adaptive	Cognitive	Social-Communication	Social	Total
Fine Motor	-----	.66* [.31 - .75]	.69* [.35 - .77]	.69* [.31 - .75]	.60* [.16 - .68]	.63* [.20 - .70]	.77* [.55 - .86]
Gross Motor		-----	.24 [-.09 - .52]	.34* [.02 - .60]	.24 [-.09 - .52]	.42** [.12 - .65]	.50** [.21 - .71]
Adaptive			-----	.47** [.18 - .69]	.22 [-.11 - .50]	.35* [.03 - .60]	.58** [.32 - .76]
Cognitive				-----	.59** [.33 - .77]	.54** [.27 - .73]	.93** [.87 - .96]
Social-Communication					-----	.56** [.29 - .75]	.75** [.57 - .86]
Social						-----	.72** [.52 - .85]
Total							-----

*p<.05, two-tailed test
**p<.01, two-tailed test

TABLE 8. Intercorrelations Among Domain and Total Test Scores for the Non-disabled Group: 95% Confidence Intervals in Bracket (N=38)

Domains/ Total	Fine Motor	Gross Motor	Adaptive	Cognitive	Social-Communication	Social	Total
Fine Motor	-----	.66* [.31 - .75]	.69* [.35 - .77]	.69* [.31 - .75]	.60* [.16 - .68]	.63* [.20 - .70]	.77* [.55 - .86]
Gross Motor		-----	.24 [-.09 - .52]	.34* [.02 - .60]	.24 [-.09 - .52]	.42** [.12 - .65]	.50** [.21 - .71]
Adaptive			-----	.47** [.18 - .69]	.22 [-.11 - .50]	.35* [.03 - .60]	.58** [.32 - .76]
Cognitive				-----	.59** [.33 - .77]	.54** [.27 - .73]	.93** [.87 - .96]
Social-Communication					-----	.56** [.29 - .75]	.75** [.57 - .86]
Social						-----	.72** [.52 - .85]
Total							-----

*p<.05, two-tailed test
**p<.01, two-tailed test

TABLE 9. Intercorrelations Among Domain and Total Test Scores for the Disabled Group: 95% Confidence Intervals in Brackets (N=12)

Domains/ Total	Fine Motor	Gross Motor	Adaptive	Cognitive	Social-Communication	Social	Total
Fine Motor	-----	.58* [.01 - .87]	.43 [-.19 - .81]	.68* [.17 - .90]	.65* [.12 - .89]	.62* [.07 - .88]	.74** [.29 - .92]
Gross Motor		-----	.51 [-.09 - .84]	.66* [.14 - .89]	.76** [.33 - .93]	.63* [.09 - .88]	.81** [.44 - .94]
Adaptive			-----	.41 [-.21 - .80]	.36 [-.27 - .77]	.53 [-.06 - .85]	.53 [-.06 - .85]
Cognitive				-----	.81** [.44 - .94]	.87** [.59 - .96]	.92** [.70 - .97]
Social-Communication					-----	.72** [.25 - .92]	.95** [.83 - .99]
Social						-----	.86** [.56 - .96]
Total							-----

*p<.05, two-tailed test
**p<.01, two-tailed test

correlated very highly with the Social-Communication (r = .95), Cognitive (r = .92), and Social (r = .86) domains. In terms of intercorrelations among domains, high correlations are observed between the Social and Cognitive domains (r = .87) and the Social-Communication and Cognitive domains (r = .81). The Adaptive domain scores for the disabled group were not significantly correlated with any domain score although the correlations are between moderate and low (r = .36 - .53). Only 12 children in the disabled group had scores in all six domains; small number of subjects may have contributed to nonsignificant correlations obtained for this group.

A high correlation between individual domains suggests the two domains contain items of similar or related construct, such as the skills measured by the Social-Communication and Cognitive domains. A low correlation indicates the two domains might measure unrelated skills, such as the skills tapped by the Adaptive and Gross Motor domains.

Construct Validity

The AEPS Test was designed to measure the developmental function of young children with delays and disabilities. It was hypothesized that the AEPS Test would be sensitive to variations in the performance of children of different ages and having different levels of disabilities. It was further

hypothesized that children who had higher ability (e.g., children who were older and children with no or less serious disabilities) would have higher domain and total test scores than children who had lower ability (e.g., children who were younger and children who had more serious disabilities). Specifically two questions were posed: What is the relationship between children's ages given no disability and performance on the AEPS Test? What is the relationship between children's degree of impairment and performance on the AEPS Test?

Subjects' ages were determined by the difference between their date of birth and their first testing date. Three age groups, three, four, and five years were formed. Subjects' degree of impairment was determined by their classroom teachers' rating on the Degree of Impairment Definitions. Five levels of impairment, non-disabled, at-risk, mild, moderate, and severe, were formed. In analyzing the test's ability to distinguish children at different ages, only children without disabilities were included.

For the analysis of the relationship between test scores and degree of impairment, the limited number of subjects required collapsing the five levels into three groups: (a) a non-disabled group composed of children with no disabilities, (b) a less impaired group composed of children who were at risk and children who had mild

disabilities, and (c) a more impaired group composed of children who had moderate and severe disabilities.

Sensitivity to Age

Data for the 62 children exhibiting no disabilities at the time the research was conducted were used to examine the extent to which the AEPS Test domain and total test scores were influenced by children's age. Three age groups were formed: (a) three-year-olds, (b) four-year-olds, and (c) five-year-olds. Four of the 62 subjects were six-year-olds; however, because they were six years two months and younger (two 74 months, one 73 months, and one 72 months), their scores were included in the five year old group. Data for subjects who were at risk and subjects with disabilities were not included in this analysis. Table 10 presents the means and standard deviations for the six domains and total test scores for the three age groups.

A one-way analysis of variance (one-way ANOVA) was used in analyzing the data for scores in each of the six domains and the total test score. A multiple comparison procedure, the Scheffé's test, was performed when a significant F was obtained. Table 11 reports the ANOVA results for the six domain and total test scores of the three age groups.

With the exception of the Adaptive domain, a significant F ratio was obtained for all domains and for the

TABLE 10. Means and Standard Deviations for AEPS Domain and Total Test Scores for the Three Age Groups

	Age								
	3			4			5		
Domains	M	SD	N	M	SD	N	M	SD	N
Fine Motor	3.08	1.55	13	2.89	2.27	18	5.55	2.80	20
Gross Motor	4.23	1.96	13	5.36	2.27	14	8.00	1.71	18
Adaptive	9.60	2.27	10	9.46	2.37	13	10.40	3.72	15
Cognitive	32.31	10.20	13	31.67	11.21	18	44.20	10.24	20
Social-Communication	39.00	3.76	13	38.94	5.84	18	43.15	2.81	20
Social	8.46	2.50	13	8.67	2.87	18	12.65	5.15	20
Total	97.90	14.87	10	104.00	16.48	11	129.00	21.06	15

TABLE 11. ANOVA Results for the Domain and Total Test Scores of the Three Age Groups

Domain	SS	df	MS	F
Fine Motor				
Between Ss	81.33	2	40.66	7.35*
Within Ss	265.65	48	5.53	
Gross Motor				
Between Ss	118.12	2	59.06	15.17*
Within Ss	163.52	42	3.89	
Adaptive				
Between Ss	7.11	2	3.56	.41
Within Ss	307.23	35	8.78	
Cognitive				
Between Ss	1831.72	2	915.86	8.17*
Within Ss	5377.97	48	112.04	
Social-Communication				
Between Ss	212.66	2	106.33	5.67*
Within Ss	899.49	48	18.74	
Social				
Between Ss	201.63	2	100.82	6.72*
Within Ss	719.78	48	15.00	
Total				
Between Ss	7008.32	2	3504.61	10.59*
Within Ss	10916.90	33	330.82	

*p<.01

total test. The Scheffé's procedure then was performed to determine which of the three age groups differed. The results of the Scheffé procedure are shown in Table 12.

For the five domain scores and the total test scores, the five-year-old group was significantly higher than the three-year-old and four-year-old groups. However, the

TABLE 12. Scheffé's Test of the Domain and Total Test Score Means of the Three Age Groups

Domain	N	Mean Difference	Scheffé's F Test
Fine Motor			
Three vs. Four	31	.19	.02
Three vs. Five	33	-2.47	4.35*
Four vs. Five	38	-2.66	6.06**
Gross Motor			
Three vs. Four	27	1.53	1.10
Three vs. Five	31	-3.77	13.77**
Four vs. Five	32	-2.64	7.06**
Cognitive			
Three vs. Four	31	.64	.17
Three vs. Five	33	-11.89	4.97*
Four vs. Five	38	-12.53	6.64**
Social-Communication			
Three vs. Four	31	.06	.00
Three vs. Five	33	-4.15	3.62*
Four vs. Five	38	-4.21	4.47*
Social			
Three vs. Four	31	-.21	.01
Three vs. Five	33	-4.19	4.61*
Four vs. Five	38	-3.98	5.01*
Total			
Three vs. Four	21	-6.10	.29
Three vs. Five	25	-31.10	8.77**
Four vs. Five	26	-25.00	5.99**

*p<.05
**p<.01

four-year-olds were not significantly higher than the three-year-olds.

Sensitivity to Degree of Impairment

The AEPS domain scores and total test scores for all 82 subjects were used in the analysis of the sensitivity of the instrument to children with different levels of impairments regardless of age. Because a number of children with disabilities were not involved in activities that assessed gross motor and adaptive behaviors, total score for this analysis is represented by the sum of four domain scores: Fine Motor, Cognitive, Social, and Social-Communication. Table 13 presents the means and standard deviations for the six domains and total test scores for the total group and two subgroups. A one-way analysis of variance (one-way ANOVA) was performed. The results are presented in Table 14. There is a statistically significant difference among the three degrees of impairment in all the domain scores and the total test scores.

The Scheffé's multiple comparison procedure then was performed to determine which of the three impairment groups were different. The results are presented in Table 15.

Table 15 shows that the children with no disabilities had higher domain and total test scores than the children with disabilities in the present sample. With the exception

TABLE 13. Means and Standard Deviations for the AEPS Domain and Total Test Scores for the Non-Disabled, At-Risk/Mild, and Moderate/Severe Groups

	Group								
	Non-Disabled			At-Risk/Mild			Moderate/Severe		
Domains	M	SD	N	M	SD	N	M	SD	N
Fine Motor	4.45	2.56	51	3.00	1.56	19	1.83	1.85	12
Gross Motor	6.24	2.52	45	3.82	1.94	11	2.71	2.56	7
Adaptive	8.95	2.88	38	5.90	3.54	10	4.91	1.76	11
Cognitive	36.75	12.01	51	16.68	7.30	19	8.58	8.45	12
Social-Communication	40.37	4.60	51	29.84	8.11	19	15.75	13.08	12
Social	8.96	3.29	51	5.26	2.35	19	3.33	2.10	12
Total[a]	90.53	18.91	51	54.84	16.55	19	29.50	23.93	12

[a]Gross Motor and Adaptive scores were not included.

TABLE 14. ANOVA Results for the Domain and Total Test Scores for the Non-Disabled, At-Risk/Mild, and Moderate/Severe Groups

Domain	SS	df	MS	F
Fine Motor				
Between Ss	79.80	2	39.90	7.68*
Within Ss	410.29	79	5.19	
Gross Motor				
Between Ss	110.05	2	55.03	9.24*
Within Ss	357.38	60	5.96	
Adaptive				
Between Ss	177.18	2	88.59	11.03*
Within Ss	449.70	56	8.03	
Cognitive				
Between Ss	10857.24	2	5428.62	47.90*
Within Ss	8952.71	79	113.33	
Social-Communication				
Between Ss	6387.41	2	3193.71	61.20*
Within Ss	4122.70	79	52.19	
Social				
Between Ss	408.28	2	204.14	23.36*
Within Ss	690.27	79	8.74	
Total				
Between Ss	44634.88	2	22316.94	60.57*
Within Ss	29106.23	79	368.43	

*p<.001

of the Fine Motor domain, significant score differences exist between children with no disabilities and the two groups of children with disabilities. Non-disabled children's Fine Motor scores are significantly higher than those of the children with more serious disabilities,

TABLE 15. Scheffé's Test of the Domain and Total Test Score Means of the Non-Disabled, At-Risk/Mild, and Moderate/Severe Groups

Domain	N	Mean Difference	Scheffé's F Test
Fine Motor			
Non-disabled vs. At-risk/Mild	70	1.45	2.81
Non-disabled vs. Mod./Severe	63	2.62	6.41**
At-risk/Mild vs. Mod./Severe	31	1.17	.96
Gross Motor			
Non-disabled vs. At-risk/Mild	56	2.43	4.37*
Non-disabled vs. Mod./Severe	52	3.53	6.33**
At-risk/Mild vs. Mod./Severe	18	1.10	.44
Adaptive			
Non-disabled vs. At-risk/Mild	48	3.05	4.58*
Non-disabled vs. Mod./Severe	49	4.04	8.66**
At-risk/Mild vs. Mod./Severe	21	.99	.32
Cognitive			
Non-disabled vs. At-risk/Mild	70	20.06	24.58**
Non-disabled vs. Mod./Severe	63	28.16	33.99**
At-risk/Mild vs. Mod./Severe	31	8.10	2.13

TABLE 15. Continued

Domain	N	Mean Difference	Scheffé's F Test
Social-Communication			
Non-disabled vs. At-risk/Mild	70	10.53	14.71**
Non-disabled vs. Mod./Severe	63	24.62	56.43**
At-risk/Mild vs. Mod./Severe	31	14.09	13.99**
Social			
Non-disabled vs. At-risk/Mild	70	3.70	10.83**
Non-disabled vs. Mod./Severe	63	5.63	17.60**
At-risk/Mild vs. Mod./Severe	31	1.93	1.57
Total[a]			
Non-disabled vs. At-risk/Mild	70	35.69	23.93**
Non-disabled vs. Mod./Severe	63	61.03	49.10**
At-risk/Mild vs. Mod./Severe	31	25.34	6.41**

*p<.05
**p<.01
[a]Gross Motor and Adaptive scores were not included

however, the difference between the non-disabled children and the children with milder disabilities was not significant.

Children with milder disabilities had significantly higher total test scores than the children with more serious disabilities. However, at the domain level, differences were significant only for the Social-Communication domain. The differences between the two groups of children with disabilities were not significant for the other five domains.

The AEPS total test scores and the six domain scores appear to be influenced by the presence and absence of disabilities in children in the sample. In addition, the total test scores seem to have the capability of distinguishing children who were at risk and with mild disabilities from those with moderate and severe disabilities. However, it should be noted that the total score in this analysis did not include Gross Motor and Adaptive scores.