Psychometric Properties of the AEPS

INTRODUCTION

The forerunner to the AEPS Test was the Adaptive Performance Instrument (API). Work began on the API in the early 1970s by a group of investigators identified as the Consortium on Adaptive Performance Evaluation (CAPE, 1978). Although many individuals were associated with CAPE and participated in the early development and testing of the API, Dale Gentry, Owen White, Lizbeth Vincent, Evelyn Lynch, Jeff Seibert, Verna Hart, Katie McCarton, and Diane Bricker provided the sustaining leadership for this project from the mid-1970s to the early 1980s (Bricker, 1981).

Initially, the AEPS was created for the developmental age range of birth to 3 years and, therefore, the initial work beginning in the late 1970s was focused on the AEPS Birth to Three Test. Development began on the AEPS Three to Six Years in the mid-1980s and the first formal study of its psychometric properties was initiated in 1985. This appendix is composed of two sections: the first section reviews the work focused on describing findings pertinent to Birth to Three Years, whereas the second section addresses the research relevant to Three to Six Years.

BIRTH TO THREE YEARS

An extensive investigation of the AEPS Test for Birth to Three Years was conducted from 1984 to 1987 with support provided by the U.S. Department of Education. The goal of the research was to provide information on selected psychometric properties of the AEPS Test (then called the EPS). Results from this project can be found in Bailey and Bricker (1986) and Bricker, Bailey, and Slentz (1990).

The overall findings from Bailey and Bricker (1986) indicated that *interrater agreement* was adequate to good for all areas of the AEPS Test for Birth to Three Years except for the Cognitive Area. *Test–retest reliability* coefficients also ranged from adequate to good for all areas. *Concurrent validity* between children's performances on the Gesell Developmental Schedule (Knobloch, Stevens, & Malone, 1980) and the AEPS Test suggested that the AEPS Test was a valid measure of children's skills and abilities for this sample of 30 children. The relationships between area scores and between area scores and total test scores were also examined. All correlations indicated a relationship between the areas and whole test but not between the individual areas. Finally, data were collected from a small sample of interventionists regarding the *utility* of the AEPS Test. These findings indicated that the AEPS Test could be successfully administered in a reasonable amount of time. Findings from the Bailey and Bricker (1986) investigation were used to make modifications to AEPS Test items.

An investigation by Bricker et al. (1990) produced similar findings regarding the AEPS Test for Birth to Three Years. First, *interrater agreement* was again found to be adequate to good for all areas (i.e., the mean correlations across areas was $r = .87$, $p<.001$). Second, test-retest reliability coefficients ranged from $r = .77$ for the Social Area to $r = .95$ for the Gross Motor Area. The

mean correlations across areas was $r = .88$ ($p<.001$). Third, concurrent validity between children's performances on the Gesell Developmental Schedule (Knobloch et al., 1980), the *Bayley Scales of Infant Development* (Bayley, 1969), and the AEPS Test suggested that the AEPS Test was a valid measure of children's skills and abilities for this sample of 155 children. Fourth, the relationship between area scores and total test scores was examined. All correlations were significant at $p<.001$.

An *item analysis* was completed by examining AEPS Test performance scores of 77 children who were typically developing and at risk ranging in age from 2 months to 4 years. The children were grouped into four categories by age: birth to 1 year, 1–2 years, 2–3 years, and 3–4 years. Items were analyzed for each age group by calculating the frequency and percent of children who scored "pass" (i.e., 2), "inconsistent" (i.e., 1), and "fail" (i.e., 0). For this sample of children, the results indicate that all items are hierarchically arranged by age. For all AEPS Test items, the percentage of items passed by older children exceeded that of younger children. When examining the hierarchical arrangement of objectives to goals, the results were also encouraging. In the birth to 1 year age group, data indicate that 81% of the item sequences are hierarchically arranged (i.e., more infants passed the easiest objectives than the more difficult ones). For the 1–2 year age group, 89% of item sequences are hierarchically arranged, whereas 90% of item sequences are hierarchically arranged for the 2–3 year age group. The 3–4 year age range was not included in the analysis because of the limited number of children who were at risk and who did not have disabilities in this group.

A similar analysis by area found only two Fine Motor, four Gross Motor, five Social-Communication, five Cognitive, three Adaptive, and six Social item sequences that were not hierarchically arranged. These data strongly suggest that for this population of children, most AEPS Test goals/objectives (80%–90%) are hierarchically arranged from simple to more complex.

Questionnaires were also completed by 23 field test sites providing feedback on the utility of the AEPS Test. The majority of sites responded favorably to the test content and the administration guide. Questions concerning test content (e.g., Do the results from the AEPS Test accurately reflect the performance of your children?) elicited "yes" responses from 76% of the field test sites—indicating that a majority of these users generally found the content of the AEPS Test appropriate and useful for developing intervention plans. Questions about the AEPS Test items (e.g., Were the item criteria clear?) were rated "yes" by 87% of the sites. Several sites commented on the usefulness of the AEPS Test in relation to the populations that they served. Questions concerning the administration guide (e.g., Did you understand how to use the scoring system?) were answered "yes" by 92% of the sites. Data indicate that the administration guide effectively defined the organizational structure and the procedures for administration of the AEPS. The amount of time it took AEPS Test users to administer the test to one child ranged from 45 minutes to 5 hours, with a mean of 2 hours and 29 minutes. Ninety-one percent of the field test sites used the AEPS Test IFSP/IEP Goal and Objective Examples. The majority of AEPS Test users (78%) indicated that the goals/objectives covered the most important behaviors to be targeted for the children that they served, as

well as saved them time in writing individualized family service plans (IFSPs)/ individualized education programs (IEPs). The amount of time saved writing IFSPs/IEPs per child ranged from 30 minutes to 2 hours, with a mean of 57 minutes saved per child.

In 1990, Notari and Bricker conducted the first in a series of studies examining the *treatment validity* of the AEPS. *Treatment validity* is defined as " . . . the meaning, relevance, and utility of scores, the import or value implications of scores as a basis for action, and the functional worth of scores in terms of the social consequences of their use" (Messick, 1989, p. 5). Treatment validity studies are necessary to determine if measures such as the AEPS Test improve the quality of services delivered to children and that, in turn, produce measurable child performance outcomes that surpass expectations or benchmarks. The primary purpose of the Notari and Bricker study was to examine the quality IFSP/IEP goals/objectives written from AEPS Test Birth to Three results. This study involved 48 early interventionists from three states and British Columbia and examined the effectiveness of the AEPS Test to generate goals/objectives that were functional, generic, measurable, and easy to embed in daily activities. The study also investigated the hierarchical relationship between goals/objectives.

The experimental groups who used the AEPS Test wrote higher quality goals/objectives than a comparison group who used other assessment tools to develop goals/objectives. The AEPS Test presented particular advantages in generating quality goals and providing guidance in the identification of specific long-term expectations for children that were developmentally and educationally consistent with the objectives.

Notari and Drinkwater (1991) investigated the quality of IFSP/IEP goals/ objectives written using the AEPS Test for Birth to Three Years. Goals and objectives for 17 children ages 3–26 months with Down syndrome were examined. In all, a total of 376 goals and objectives were rated with 224 from the AEPS Test and 152 from a computerized list (i.e., a list compiled by educators and therapists). Sets of goals/objectives were randomly combined and rated. An independent sample t-test was performed to determine whether significant differences existed in the quality of child IFSP/IEP goals/objectives and the hierarchical relationship between them. Significant differences were found between the two groups' goals/objectives, suggesting that when teachers used the AEPS Test for Birth to Three Years, they wrote goals/objectives that were "more functional, generic, easy to integrate within the instruc-tional content, and measurable, as compared to those based on computerized list" (p. 101). Notari and Drinkwater also reported that "the AEPS enabled teachers to sequence goals/objectives according to a hierarchical teaching sequence" (p. 101).

A study conducted by Cripe (1990) examined the *effectiveness of the AEPS family measures* (e.g., Family Interest Survey and Family Report, then called the Parent Form) and specific training on how to use the AEPS linked assessment, intervention, and evaluation system. Thirty interventionists from three states were divided into three treatment groups. The pretest measure was an IFSP completed on a child and family by the interventionists within the previous 12 months. The posttest measure was a second IFSP written following treatment.

Group 1 received the AEPS family measures, Group 2 received the same materials plus a half-day of training, and Group 3 received the same materials plus 2 days of training. The results indicated that interventionists in all groups included significantly more family-related outcomes on the IFSP at posttest. In addition, the quality of ratings (e.g., measurability, specificity, functionality, generality) was higher for IFSP outcomes at posttest for all groups. Group three, who received the materials plus 2 days of training, demonstrated the greatest improvement.

Straka (1994) conducted an investigation that compared the utility of outcomes of the AEPS Test with the Communication Symbolic and Behavior Scales (CSBS; Wetherby & Prizant, 1993) in regards to intervention planning for young children with communication problems. Four American Speech-Language-Hearing Association certified communication specialists participated in the study. Straka (1994) reported that communication specialists found the AEPS Test to be more useful than the CSBS in developing IEP goals/objectives. Straka reported few differences, however, in functionality, generality, instructional context, and measurability of goals/objectives developed from the CSBS and AEPS. Findings also indicated that the AEPS Test provided an easier format for developing goals/objectives (more efficient and less time consuming) and more guidance in identifying specific long-term goals/objectives. Lastly, Straka reported that the communication specialists found the AEPS Test to be more helpful in developing functional and appropriate goals/objectives that corresponded to the individualized needs of each child.

In 1995, Hamilton continued the investigation of the *treatment validity* of the AEPS Test by examining the quality of educational goals/objectives for young children with visual impairments. Forty special education teachers served as subjects. All teachers received 1 day of training on using the AEPS Test to write goals/objectives. Hamilton (1995) found that goals/objectives written by teachers for children with visual impairments were of higher quality when they used the AEPS Test versus the Oregon Project Curriculum for Visually Impaired and Blind Preschool Children (Anderson, Boigon, & Davis, 1991). Test main effect was significant, $F(1, 37) = 20.66$, $p<.001$. Teacher-written goals/objectives for young children with visual impairments based on AEPS Test information ($M = 21.20$) were of significantly higher quality overall, than goals/objectives based on Oregon Project information ($M = 15.95$).

THREE TO SIX YEARS

As indicated in the previous section, psychometric investigations of the AEPS Test for Birth to Three Years began in the late 1970s, whereas psychometric investigations of the AEPS Test for Three to Six Years began in the mid-1980s. The first formal study was completed in 1986.

The 1986 Slentz study examined selected psychometric properties of the AEPS Test for Three to Six Years. Fifty-four children who ranged in age from 34 to 76 months participated in the study. Seventeen children showed no delays in development, 22 were at risk for delays, and 15 were identified as having disabilities: 7 mild and 8 with more moderate disabilities. *Interobserver*

agreement was found to be adequate to good for all areas and significant at $p<.001$ (i.e., correlations ranged from $r = .60$ to $r = .94$). *Test–retest reliability* coefficients ranged from adequate to good for all areas except the Gross Motor and Adaptive Areas. *Concurrent validity* between children's performances on McCarthy Scales of Children's Abilities (McCarthy, 1972), the Uniform Performance Assessment System (Haring, White, Edgar, Affleck, & Hayden, 1981), and the AEPS Test suggested that the AEPS Test was a valid measure of children's skills and abilities for a sub-sample of 18 children. The relationship between area scores and total test scores were also examined. All correlations were significant and ranged from $r = .37$ ($p<.01$) for the Adaptive Area to .97 ($p<.0001$) for the Cognitive Area.

The mean rank order of AEPS Test total scores was calculated for 3-year-olds in comparison with 4-year-olds and for 4-year-olds in comparison with 5-year-olds. A Mann-Whitney U test found that 4-year-olds' total test scores were significantly higher than 3-year-olds' total test scores ($p<.001$). The difference between 4- and 5-year-olds' total test scores was not significant—likely because of the small sample and the children's location on the age continuum (i.e., many 4-year-olds were close to 48 months of age).

Psychometric properties of the AEPS Test for Three to Six Years were also studied by Hsia (1993). Eighty-two children who ranged in age from 3 to 6 years participated in this study. Fifty-one of these children were classified as having no disabilities, 20 were classified as having disabilities (i.e., had been determined eligible to receive early intervention services), and 11 were classified as at risk. An AEPS Test was completed on each child by a trained examiner. The test was scored primarily through observation of the children as they participated in planned and free play activities while attending a center-based program.

Interobserver reliability was examined by correlating the children's area scores and total test scores from the AEPS Test protocols of two independent coders. Data from all 82 children were included in this analysis, although not all children had the opportunity to participate in all gross motor and adaptive activities necessary for scoring a complete protocol. The percent agreement between two independent coders ranged from .86 to .95 for individual areas and .90 for the total test.

A high degree of internal consistency ($r = .99$) was found for the total group and for the two subgroups (children with or who are at risk for disabilities, $r = .97$; children without disabilities, $r = .93$). At the area level, the total group correlations were .75 for the Fine Motor Area, .82 for the Adaptive Area, .83 for the Gross Motor and Social Areas, .95 for the Cognitive Area, .96 for the Social-Communication Area, and .97 for the total area scores. All correlations were significant at the $p<.01$ level. Correlations tended to be lower for the subgroups, but all were significant except for the Social Area for the subgroup with disabilities.

Hsia (1993) also examined the *sensitivity* of the AEPS Test. The AEPS Test was designed to be sensitive to variations in the performance of children of different ages and with different levels of disability; therefore, children who are older and are more able should score higher on the test than children who are younger or less able.

In analyzing the test's accuracy at distinguishing children of different ages, Hsia (1993) included only children *without disabilities.* For this analysis, children were assigned to one of three age groups (3-, 4-, and 5-year-old groups) based on their date of birth. A one-way analysis of variance was used to examine subjects' scores in each of the six areas and for the total test score. A multiple comparison procedure, the Scheffe test, was performed when a significant *F* was obtained. With the exception of the Adaptive Area, a significant *F* ratio was obtained for the other five areas and for the total test scores. The 5-year-olds scored significantly higher than the 3- and 4-year-olds, but the 3- and 4-year-olds were not significantly different from each other.

Because a number of children with disabilities were not involved in activities that assessed their gross motor and adaptive abilities, total scores for the children with disabilities for this analysis were obtained by summing the other four area scores (i.e., Fine Motor, Cognitive, Social-Communication, and Social). A one-way analysis of variance was performed using three groups of children: those without disabilities, those at risk or with mild disabilities, and those with moderate/severe disabilities. A statistically significant difference was found for all groups for all areas and for the total test. The Scheffe test indicated that children in the group without disabilities had higher area and total test scores than the children in both of the other groups.

With the exception of the Fine Motor Area, significant score differences were found between the three groups. The Fine Motor Area scores of the group without disabilities were significantly higher than those of the group with moderate/severe disabilities but not those of the group at risk or with mild disabilities. The group at risk or with mild disabilities had significantly higher total test scores than the group with moderate/severe disabilities, but, at the area level, only Social-Communication Area scores were significantly different.

A series of studies funded by the U.S. Department of Education was conducted beginning in the mid-1990s and continued the focus on the *treatment validity* of the AEPS Test for Three to Six Years (Bricker & Pretti-Frontczak, 1997). To examine the treatment validity of the AEPS Test, a series of hypotheses were generated. First, it was hypothesized that administration and use of the AEPS Test would improve the quality of written IFSP/IEP goals/objectives. Second, it was hypothesized that better written IFSP/IEP goals/objectives would increase teachers' use of a specific intervention strategy referred to as *embedding.* Third, it was hypothesized that the embedding of quality goals/objectives during a variety of daily activities would lead to improved outcomes for young children.

Pretti-Frontczak and Bricker (2000) conducted a *treatment validity* study based on the hypothesis that using AEPS Test results produced higher quality goals/objectives. In particular, they examined whether a 2-day training session on how to write IEP goals/objectives and the use of the AEPS Test results improved the quality of IEP goals/objectives written by early childhood special education personnel. Findings are based on a comparison of pretraining goal/objective ratings with posttraining goal/objective ratings for 86 participants from five states. Participants (i.e., teachers, therapists, coordinators) for the study were somewhat geographically diverse representing the middle, south-

ern, and western regions of the country. Using paired t-tests, aggregated mean percent scores for goals/objectives using the AEPS Test before and following training were compared. Statistically significant differences for aggregated mean percent scores for each of the 10 goal quality indicator comparisons were found. In addition, statistically significant differences for aggregated mean percent scores for 9 of the 11 objective quality comparisons were found. The results of this finding are clear: use of the AEPS Test Three to Six Years in combination with training on writing goals/objectives improved the quality of goals/objectives written by participants.

In order to examine the second hypothesis associated with treatment validity (i.e., better written IFSP/IEP goals/objectives would increase teachers' use of a specific intervention strategy referred to as *embedding*), two additional studies were conducted. In the first study, 16 early childhood educators and early childhood special educators from four states participated (Bricker & Pretti-Frontczak, 1997; Pretti-Frontczak & Bricker, 2001). The teachers attended a two-day training session on how to use the AEPS Test to assess children and to write goals/objectives from assessment results. The quality of written goals/objectives, as well as the teacher's use of the embedding strategy before and after training, was measured.

Findings from the study of 16 early childhood education (ECE)/early childhood special education (ECSE) teachers indicated a significant improvement in the quality of *goals* written after training on all 10 quality indicators, which is especially noteworthy give the small sample size. Findings also indicated a significant ($p<.001$ to $p<.05$) improvement in the quality of written *objectives* on 7 of 11 quality indicators following training. A second analysis revealed that overall, 13 of the 16 participants improved the quality of objectives targeted specifically for use with the embedding strategy. Despite the writing of higher quality goals/objectives and the targeting of higher quality objectives by 13 of the 16 teachers, the percentage of intervals that teachers used the embedding strategy decreased from the pretraining phase to the posttraining phase for 11 of the 16 participants. A relationship between the decrease in teachers' use of the embedding strategy and the quality of target objectives was not found.

A follow-up study was conducted in which a group of ECE/ECSE teachers were given training on writing quality goals/objectives from AEPS Test results **and** explicit training on use of the embedding strategy across daily classroom activities (Pretti-Frontczak & Bricker, 2001). Seven ECE/ECSE teachers were recruited to participate in the follow-up study.

Findings from the follow-up study indicate that all objectives written from the AEPS Test results met the quality criteria rating of 7 or higher (i.e., 7 out of 11). Findings also indicated that the seven teachers used the embedding strategy with their target child an average of 9.7% of intervals observed on one measure and 10.6% of intervals observed on a second measure. The teachers in the study tended to embed target objectives by asking questions and providing verbal models. Lastly, the teachers were most likely to utilize the embedding strategy when they worked individually with the target children and engaged in language and pre-academic activities. The teachers were least likely to embed children's objectives during large group activities or while addressing adaptive skills.

When conducting treatment validity studies, it is important to examine outcomes that have relevance for the ultimate target—in this case—children. Two single subject studies were designed and conducted to determine if higher quality objectives written from AEPS Test results and embedded into daily activities would improve child outcomes (Bricker & Pretti-Frontczak, 1997).

In Study One, four preschool teachers from the northwestern region of the United States participated in the study. The teachers selected one child from their classroom for whom they targeted two objectives written from AEPS Test results and embedded during daily activities. Visual analysis was used to examine the functional relationship between using the embedding strategy with quality objectives written from the AEPS Test results and children's performance on targeted objectives. Child performance data from Study One suggest that, in general, a functional relationship exists between embedding children's objectives into daily activities and progress made on targeted objectives for training group children. Measures of the children's progress revealed changes from the baseline phase to the intervention phase in both trend and slope. Although the results of Study One were promising, they were mixed—not all children responded to the same extent when compared with one another's progress or when compared with their two targeted objectives. Moreover, the results were based on two targeted objectives for only four children. To further examine the relationship between teachers' use of the embedding strategy and child performance, a replication study was conducted.

The replication study was conducted with four different teachers. Each teacher selected one child from her classroom for whom she targeted two intervention objectives developed from the AEPS Test results. Teachers then were asked to embed these selected objectives in daily classroom activities. As in the previous study, visual analysis was used to examine the relationship between the embedding of quality objectives and child performance on these objectives. Although measures of child performance revealed some change from baseline to intervention in both trend and slope, the results were not compelling. Although the children did make progress toward their objectives, the results did not demonstrate a functional relationship between the teachers' use of the embedding strategy and child performance.

Results from these two studies provide limited evidence that teachers' embedding of quality objectives is associated with positive change in child performance. Continued work is needed to determine if the rate and length of time that teachers embed children's quality objectives are critical variables in effecting child performance over time.

SUMMARY

Ascertaining adequate support to conduct well-designed investigations of the AEPS Test has been difficult. To gather data from a range of subjects and interventionists from a variety of geographic locations is a costly endeavor. Although the investigations that have been conducted fall short of what should be done, the outcomes do provide some assurance that the AEPS Test is generally reliable and valid for the samples involved. We are grateful for the sup-

port provided by the U.S. Department of Education to conduct the major investigations reported in this appendix.

Perhaps the more important outcomes are those that suggest that using the AEPS Test improves the quality of IFSP/IEP goals and objectives developed by interventionists. Also, the research that has examined the utility of the test has provided extremely useful feedback. Again, this work only touches the surface of what needs to be done to establish that the AEPS Test and accompanying materials are effective and useful across a range of interventionists and children.

The developers of the AEPS hope to continue studying the properties, effectiveness, and usefulness of the materials as time and resources permit.