7

AEPS®-3 Research

Accurate and comprehensive assessment of children's skills across important developmental areas is a bedrock necessity for effective intervention. To be effective, intervention and teaching efforts have to begin by targeting the next developmental step as children move toward independence and competence. Focusing on teaching skills that children have mastered is a waste of valuable resources, as is targeting skills that are too advanced given a child's current ability. The importance of accurate assessment of children's baseline skills levels and subsequent measurement of their progress cannot be overstated. In addition, determining the usefulness and effectiveness of curricular content is of critical importance.

Using a comprehensive measure of all important areas of early development—with empirically demonstrated validity, reliability, and utility—is fundamental to accurate assessment of children's developmental skills and progress monitoring. Measures of validity indicate a measure's accuracy in evaluating the presence or absence of targeted developmental skills such as whether a child can walk using alternating steps. Important dimensions of reliability offer data on whether independent assessors score items the same way (*interrater reliability,* or *IRR*) as well as whether assessors score the measure similarly over a short time period (test–retest reliability). Measures of utility can vary widely, but important aspects that are usually addressed include determining clarity of instructions and usefulness of results.

Determining the effects of using a particular curriculum on children's development over time is more challenging but no less important. Increasingly, the field of early intervention/early childhood special education (EI/ECSE) is finding ways to assess curricular impact through careful progress monitoring using psychometrically sound measures such as AEPS®-3.

The AEPS-3 developers are committed to careful study of the validity, reliability, and utility of the AEPS-3 Test, Ready-Set, and Curriculum, as well as to the quality and usefulness of associated materials. As with previous editions, findings continue to support the utility, validity, and reliability of the AEPS-3 Test. Perhaps most importantly, AEPS-3 Test results can be used to determine eligibility for services.

This chapter provides important information on the psychometric properties of the AEPS-3 Test as well as the data collection and analyses associated with other components of the system that have been assembled to date. The chapter begins with a description of the study conducted to ensure the content validity of the revised AEPS-3 Test. The second section provides details on the field test study conducted to examine the validity and reliability of the AEPS-3 Test. This section also addresses how data were collected and analyzed to determine the AEPS-3 cutoff scores that can be used to establish eligibility for publicly funded EI/ECSE services (see also Volume 2, Chapter 8). The next section describes preliminary data collected on Ready-Set. The following section describes a preliminary validation study to examine the utility of the AEPS-3 Curriculum. The final sections briefly review development of the AEPS-3 Spanish translation, various support materials, and international research using translations of the AEPS-3 Test and support components. Interested readers can find a complete list of published research associated with AEPS at https://aeps-emrg.weebly.com.

AEPS-3 CONTENT VALIDITY STUDY

As noted in the Introduction, revisions to AEPS focused on three areas: content, organization, and format. Content changes are arguably the most important because content is the essence of a measure from which all other aspects evolve. Once all initial changes were made, the authors reviewed the revised developmental area items with particular attention given to the newly created Literacy and Math areas. All items were assigned to strands, and their goals and associated objectives were selected. Items considered too difficult, too similar, or too challenging were eliminated. Even after this process, the number of items remaining in each area was still high, and further culling and evaluation were necessary. Consequently, in 2012, the authors, with Brookes staff, designed a content validity study. Its purpose was to gather qualitative feedback from experts in early development on the items in the AEPS-3 Test's eight developmental areas (Macy, Bricker, et al., 2015).

Expert reviewers were selected for each developmental area with four or five reviewers assigned to a developmental area matching their expertise. Criteria for these reviewers included the following:

- Holding an advanced degree

- Completing relevant work in the area, such as published papers or research in early childhood, early childhood special education (ECSE), or a related field

- Meeting project timelines

A separate group of reviewers examined content from the entire test. These were individuals with broad-based knowledge of child development and intervention. Of 88 potential reviewers contacted, 37 agreed to participate and were able to meet the timeline for providing feedback. Each reviewer received a Word document of the area(s) to review and a copy of the complete list of areas, strands, goals, and objectives for the revised AEPS-3 Test. The expert reviewers provided written responses to five questions:

1. Are items in correct developmental order? Specifically, are the goals ordered developmentally within strands, and are objectives ordered developmentally within goals?

2. Are there gaps or redundancies in goal content?

3. Are there items or criteria that lack clarity?

4. Are there items that are not related to the content or skill?

5. Are all items functional, teachable skills?

The form of qualitative feedback received varied across expert reviewers. Most reviewers answered the five questions in a separate document containing their feedback and also wrote comments on the Word document for their assigned area(s). Reviewers returned feedback and forms directly to Brookes editorial staff, who removed identifying information from documents before sending them to the authors.

Eight author work groups (one for each test area) conducted the qualitative analysis. Each work group had a chairperson who first reviewed the unidentified feedback from both the expert reviewers in their area and the overall reviewers and then completed a Content Validity Analysis form using each expert reviewer's feedback. The chairperson then sent the Content Validity Analysis form to each work group member for review and ensured that all members entered their feedback. Once all members had completed the form, changes were made if work group members uniformly agreed to the expert reviewer's recommendations. When disagreements occurred between an expert reviewer and work group participants, the item was noted and scheduled for a follow-up discussion at a subsequent face-to-face meeting of all authors. At this meeting, the authors discussed the recommendations, resolved disagreements, and completed the final editing of areas.

Qualitative feedback for the content validity analyses came from two sources: the reviewers' answers to the five feedback questions listed earlier and the reviewers' additional comments. Most reviewers answered no to question 3 (Are there items or criteria that lack clarity?) and question 4 (Are there items that are not related to the content or skill?). Most reviewers answered yes to question 1 (Are items in correct developmental order?), question 2 (Are there gaps or redundancies in goal content?), and question 5 (Are all items functional, teachable skills?). The reviewers commented on gaps and redundancies that resulted from merging the two levels of the test. Expert reviewers also made written notes on the test protocol and shared feedback in Word documents. Work group members addressed all comments and documented the findings.

The authors and Brookes staff reviewed and finalized the test item content for all areas. In general, feedback on the revised test items was positive, and suggestions for content changes and reordering were mostly minor. As expected, the two new areas of Literacy and Math received the most extensive attention from the reviewers and underwent the most substantive modifications. Feedback on the Literacy area took additional time to analyze because expert opinions on the developmental content and sequence for early literacy differed (McCardle & Chhabra, 2004; National Research Council, 1998), with reviewer recommendations often opposing each other. Feedback on the *Math area* also varied considerably.

This study had two significant limitations. First, members of the author group and Brookes staff made recommendations for the expert panel, and as a result, there may have been a conscious or unconscious bias toward nominating individuals with shared values. Other experts might have offered different and possibly more negative feedback. Second, employing other more empirically based research designs and methodologies might have produced different results. Other analytical approaches, such as Rasch (Winchell, 2011) or the Delphi technique, might have yielded outcomes far different from those of these qualitative analyses. The way the feedback was collected and examined no doubt affected the outcomes.

Nevertheless, the authors found the expert reviewer feedback extremely helpful in identifying issues with the content and developmental sequencing. The resulting series of changes should enhance the validity of the AEPS Test, though subsequent investigations are needed to verify this impression. The AEPS-3 Test field test studies were designed, in part, to address this goal.

AEPS-3 TEST FIELD TEST STUDIES

Once feedback from the expert reviewer panel had been incorporated, the next step was to finalize the items for each of the eight areas. Once that task was completed in 2015, field testing began. Field testing included two studies. The first study focused on gathering information on the utility, validity, and interrater reliability of the revised test (Grisham et al., 2020). The second study collected the necessary child performance data to establish empirically derived cutoff scores and then to examine the accuracy of those cutoff scores for determining children's eligibility for services (Toland et al., 2021). The next section presents these studies' details. As noted, field test studies examined three important components of the revised AEPS-3 Test: utility, validity, and reliability.

Utility Study

Before child data were collected for the concurrent validity and interrater reliability studies, a utility study gathered information about AEPS-3 Test content and scoring to identify any problems. Subjects for this study included a small sample of field test participants from Kentucky and Oregon who assessed children using the experimental AEPS-3 field test protocol and identified concerns with test content and scoring. Participants were required to take an AEPS-3 online training before collecting AEPS-3 Test data.

The utility study's participants were 11 providers from two sites, both of which required use by staff of the second edition of the AEPS Test as their primary curriculum-based assessment. Three participants had bachelor's degrees and eight had postgraduate degrees. Most of the participants had extensive teaching experience as well as experience using previous editions of the AEPS Test. Each provider received a stipend and a set of professional resources in exchange for accomplishing several tasks: completing the AEPS-3 online training, assessing one or more children with the AEPS-3 Test protocol, and completing the utility survey.

The 11 providers assessed 23 children between the ages of 4 and 83 months, with assessment observations taking place in either the child's home or the classroom setting. After administering the AEPS-3 Test, providers completed a utility survey that asked questions designed to evaluate the revised scoring, the content of the goals in each of the eight areas, and the usefulness of the AEPS-3 Test for its intended purposes.

Survey findings indicated that participants found the majority of goals to be functional and teachable and the items and criteria easy to understand. In addition, the majority of participants either agreed or strongly agreed that the AEPS-3 Test meets its intended purposes, which include the following:

- The test can be used in authentic environments.

- The test can provide useful information for summarizing children's strengths and present levels of development.

- Information from the test can inform outcome data necessary for reporting to state and federal agencies.

- The test can monitor children's progress.

- The revised scoring can provide useful information for progress monitoring.

Based on results of the utility study, the authors revised some test item content and criteria to improve clarity and removed one item from the Literacy area.

Concurrent Validity Study

The concurrent validity study was conducted at a field test site in Kentucky with eight teacher/provider participants whose years of experience ranged from less than 1 year to 9 years. Each teacher/provider participant recruited from their own program a minimum of four children birth to 6 years old, with or without disabilities, to participate in the study. They recruited a total of 50 children (25 females and 25 males) between the ages of 12 months and 65 months. The children's ethnicities were as follows:

- Caucasian (68%)

- African American (4%)

- Hispanic (4%)

- Asian (14%)

- Other (2%)

- Mixed (8%)

Of the 50 children, 10 were receiving special education services and 40 were not receiving services.

The AEPS-3 Test's concurrent validity was examined by completing, on the same children within 2 weeks, both the AEPS-3 Test and the Battelle Developmental Inventory–Second Edition, or BDI-2 (Newborg, 2005). Teachers/providers completed the AEPS-3 Test by observing children in their classrooms. The teachers/providers had received previous training on administering the BDI-2, all had used the AEPS Test multiple times, and all had completed the AEPS-3 Test online training module and interrater reliability test before completing the AEPS-3 Test. Participating providers were required to score and enter all items on the AEPS-3 Test using the Child Observation Data Form (CODF). BDI-2 data were collected within 2 weeks of completing the AEPS-3 Test.

Concurrent validity was examined using correlational comparison of children's scores on the BDI-2 and the AEPS-3 Test. Pearson correlations were calculated in two ways:

1. The first analysis calculated a total AEPS-3 Test score for each area by combining scores for goals and objectives and then compared these scores with area BDI-2 scores.

2. The second analysis compared AEPS-3 total goal scores for each area only—that is, scores were calculated without the objectives—with BDI-2 area scores.

The results indicated positive correlations in the weak to moderate range (r = .31 to .65) for AEPS-3 combined goal area scores with BDI-2 area scores. All correlations were statistically significant except for the AEPS-3 Fine Motor combined area score and the BDI-2 Adaptive area score (r = .24).

As might be expected, larger positive correlation coefficients occurred when comparing similar areas across the two tests. For example, the correlation between AEPS-3 Test's Social-Communication scores and the BDI-2 Communication scores was positive (r = .63), whereas a less robust correlation (r = .35) occurred when comparing the AEPS-3 Test's Social-Communication scores and BDI-2 Motor scores.

When using the AEPS-3 Test's goals-only area scores, Pearson correlation coefficients for the BDI-2 area scores were lower. Correlations across areas were consistently higher when using AEPS-3 Test combined goals and objectives scores. The results also indicated a stronger correlation between age and AEPS-3 Test combined scores (r = .65 to .92) than between age and AEPS-3 Test area goal scores only (r = .34 to .84).

Interrater Reliability Study

A total of 116 providers (115 females and 1 male) from the 14 field test sites participated in the interrater reliability study. Providers came from seven states: Kansas, Kentucky, Ohio, Oregon, Tennessee, Texas, and Virginia. More than 90% of the participating providers had a bachelor's or postgraduate/graduate degree. The average number of years of experience for participants working with children was 6+ (range = 0 to 34).

Prior to completing the interrater reliability test, all teacher/provider participants were required to first view a 1-hour online training module that described the differences between the second and third editions of the AEPS Test and explained the new scoring rules and guidelines for AEPS-3. The training module included embedded video clips that allowed participants to practice scoring AEPS-3 Test items.

Once teachers/providers completed the training module, they participated in the interrater reliability test that consisted of 37 video clips of children engaged in typical activities in home and classroom environments. The clips included scoring observations for 68 AEPS-3 Test items taken from all eight areas of the test and across all developmental levels. Participants watched video clips and then scored the items addressed in each clip using the AEPS-3 Test 2, 1, 0 scoring criteria. After scoring all the videos, participants received immediate feedback on their results.

Participant responses were compared with predetermined expert scoring of items on each video clip. Provider interrater agreement with AEPS-3 authors/experts ranged from 66% to 100%, with a grand mean of 89%. These results suggest that participants with experience using the AEPS Test can reliably score children's performance on the AEPS-3 Test after completing a brief online AEPS-3 scoring training module. Study details can be found in Grisham et al. (2020).

AEPS-3 TEST CUTOFF SCORE DETERMINATION

The first part of the field study collected child performance data to examine the utility, validity, and reliability of the AEPS-3 Test. The second phase involved using this data set to conduct an analysis to establish empirically derived cutoff scores that could be used to determine eligibility for Part B and Part C Individuals with Disabilities Education Improvement Act (IDEIA) services (Toland et al., 2021).

Subjects included 874 children, of whom 47.1% ($n = 412$) were children without disabilities (ineligible for services) and 52.9% ($n = 462$) were children receiving IDEA services (eligible for services). Children ranged in chronological age from 2 months to 6 years 11 months and included 515 boys and 359 girls. Children were recruited in seven states by programs currently using the AEPS Test. Program types included home visiting, parent–toddler groups, child care centers, publicly funded prekindergarten programs, and Head Start programs.

Prior to collecting data on children, teachers had to participate in the online AEPS-3 training and interrater reliability test (described above). Those who received a score of 80% or higher on the interrater reliability test were allowed to move forward with data collection.

Teachers/providers collected observational data on children and then used this information to score items based on AEPS-3 Test written item criteria. Teachers/providers entered children's item scores into a secure online data entry web site, using a unique code assigned to each child participant to ensure confidentiality. Children's performance data were then used to examine the AEPS-3 Test's psychometric dimensions and develop cutoff scores.

The quality of AEPS-3 Test items was addressed by conducting a unidimensional Rasch rating scale model (RSM) analysis (cf. Andrich, 1978). An underlying assumption of the RSM is unidimensionality. Toland et al. (2021) provide a detailed description of the process and steps used to determine unidimensionality. When the analysis identified concerns about dimensionality, a further inspection of the items was conducted. To determine whether items were performing in a manner consistent with the RSM, infit and outfit item indices were assessed, and those with unacceptable values were considered for removal, one item at a time (iteratively), until all items had an acceptable fit to RSM. An item was not removed if its removal jeopardized content coverage for an area, if the degree of misfit was considered tolerable (close to the expected infit and outfit range), or if the item was deemed important for programming purposes.

In addition, person reliability (the degree to which an area test separates a relevant child into enough levels) and item reliability (a measure that indicates item spread along a particular area test continuum) were examined. Then, once items were calibrated, an analysis was conducted to create cutoff scores for each developmental area at 6-month age intervals (an interval used to ensure stability).

Once cutoff scores had been determined, they were used to establish eligibility classifications for the entire child sample. Specifically, children were classified as eligible if they had at least two AEPS-3 Test area scores at or below the cutoff scores for their age interval. Otherwise, they were classified as ineligible.

To examine the validity of the cutoff scores from the AEPS-3 Test, sensitivity and specificity were calculated by using cross-tabulation between known eligibility status (provided before the data analysis) and identified eligibility classification based on the AEPS-3 Test cutoff scores. Analyses made it possible to establish valid 6-month age interval cutoff scores that could be used to accurately classify children as eligible or ineligible for services. Overall, sensitivity ranged from a low of 57% at the 61- to

66-month interval to a high of 100% at the 0–6-month and 7–12-month age intervals. Specificity ranged from a low of 0% at the 0–6-month and 7–12-month age intervals to a high of 81% at the 67–72-month age interval. False-positive rates (ranging from 19% to 87%) were generally higher than false-negative rates (ranging from 6% to 43%). Nominal agreement (accuracy) ranged from 31% to 84%.

Sensitivity and specificity also were examined using a receiver operator characteristic (ROC) curve analysis to test for statistical significance. The area under the curve (AUC) values are tested for statistical significance represented by five ratings: excellent (between 0.9 and 1), good (between 0.8 and 0.9), fair (between 0.7 and 0.8), poor (0.6 and 0.7), and failed (0.5 and 0.6) (Metz, 1978). The AUC values were significant and considered good to fair for most age intervals except for the two youngest age intervals (0–6 months and 7–12 months), which were not significant and were considered poor. It is important to note that the two bottom age intervals had small sample sizes: $N = 16$ for the 0–6-month age interval and $N = 39$ for the 7–12-month interval, which could be a factor in the lack of statistical significance and low value rating.

In summary, AEPS-3 cutoff scores for the field test version identified a high percentage of children who were eligible for services (57%–100%) and a moderate to higher percentage of children who were not eligible for services (0%–81%). The rates for sensitivity and specificity for the 0- to 6-month and 7- to 12-month age intervals are least robust, suggesting that caution should be used when interpreting them because of the small sample size and the inability of the cutoff scores to accurately classify children as ineligible. In addition, results from the reliability assessment measure suggest that each developmental area may be used in isolation for other educational and clinical use.

Bricker, Clifford, et al. (2008) found that AEPS 2nd edition effectively corroborated eligibility decisions. Similar to that study, findings from the present study suggest that AEPS-3 cutoff scores are not likely to misidentify children eligible for services but cutoff scores may tend to overidentify children who do not need services.

See Toland et al. (2021) for the results.

Field Test Findings Summary

In general, the findings from the current investigations support research conducted on previous versions of the AEPS Test. The field test version of AEPS-3 provides useful information about young children's development and is useful in making a variety of teaching and intervention decisions. Results from this study further indicate that practitioners may use AEPS for eligibility determination. Complete details about this study can be found in Toland et al. (2021).

AEPS-3 READY-SET STUDIES

Ready-Set is a brief assessment and progress monitoring measure that uses advanced items from the AEPS-3 Test to target skills children need to be successful in kindergarten and first grade. This new AEPS-3 component offers educational programs a way to efficiently and accurately measure the development of 4- and 5-year-olds who may be typically developing or at risk for developmental delays. Ready-Set also can be used with older preschool children with mild disabilities to determine relevant teaching/intervention content. Volume 2, Chapter 6, offers a more detailed description of Ready-Set. Evaluation of this new component is ongoing. The next sections describe two preliminary studies.

Relationship Between Teacher and Parent Evaluations Study

This study examined use of AEPS-3 Ready-Set with a population of 4- to 6-year-old children who attended Montessori schools in Florida and Idaho (Macy, Pool, et al., 2021). The study addressed two questions:

1. What is the relationship between teacher-completed Ready-Set scores and parent-completed Ready-Set Family Assessment of Child Skills (FACS) scores?

2. Do teachers and assistants find completing Ready-Set useful?

The study included three types of participants associated with Montessori preschool programs located in either Florida or Idaho: 37 children ages 4 to 6 years, their parents, and Montessori teachers and assistants ($N = 32$). Data collection involved using three components: the AEPS-3 Ready-Set, the AEPS-3 Ready-Set FACS (which contains the same items that appear on Ready-Set but written using family-friendly wording with examples), and the utility survey (a 22-item utility survey developed to

obtain feedback from teachers and assistants on the usefulness of Ready-Set results). Survey questions focused on 1) clarity of Ready-Set items, 2) usefulness of Ready-Set in providing information about a child's current skills, 3) whether Ready-Set results provide information for summarizing individual child strengths related to the child's readiness skills for kindergarten, and 4) whether Ready-Set yields meaningful information for parent–teacher meetings. Teachers could score items by using a four-point Likert scale (Strongly Agree, Agree, Disagree, Strongly Disagree), by responding yes or no, or by selecting from multiple choice options. They also could leave comments on individual survey questions and prompts.

The study involved four steps. First, parents and teachers/assistants signed letters of informed consent. To participate, the teachers/assistants had to have in their class preschool-age children scheduled to enter kindergarten the next year. Second, teachers/assistants received instruction and a 2-hour training from research study personnel on how to accurately complete Ready-Set for participating children and how to help participating parents accurately complete the Ready-Set FACS. Third, teachers/assistants independently completed Ready-Set for each child. During the same time period, parents completed a Ready-Set FACS. Fourth, teachers/assistants completed a utility survey on the Ready-Set results.

Agreement between teacher/assistant-completed Ready-Set scores and parent-completed Ready-Set FACS scores was examined. Scores were calculated by adding the number of points for each item (Ready-Set FACS response point values were 2 for yes, 1 for sometimes, and 0 for not yet). Correlations were obtained for all eight developmental area scores and total scores for the two measures (i.e., AEPS-3 Ready-Set and Ready-Set FACS). Positive correlation coefficients were found in most areas, suggesting agreement between teacher/assistant and parent evaluation of child skills across the two measures.

Feedback gathered from the utility survey suggested that teachers/assistants found Ready-Set to be useful. For example, 97% of the teachers/assistants reported that Ready-Set results were useful for communicating child status to parents. For more information about this study, see Macy, Pool, et al. (2021).

Content Validity and Utility Study

This dissertation study posed two important questions about features of AEPS-3 Ready-Set (Stevenson, 2019). The first question addressed the content validity of AEPS-3 Ready-Set by asking 36 kindergarten teachers to rate the relevancy, breadth, and clarity of all 40 Ready-Set items using an online Qualtrics survey. Participants reviewed the AEPS-3 Ready-Set items and then completed the survey, designed to evaluate their relevancy, breadth, and clarity. Teacher responses regarding Ready-Set item relatedness, item clarity, and item criterion were analyzed. Participants reported that the Ready-Set items were related according to developmental area and that items and criteria were generally clear. Reported scale-level content validity index based on the average (S-CVI/AVE) calculations were 96% for item relatedness, 96% for item clarity, and 97% for criterion clarity.

The second question evaluated the utility of Ready-Set for use with young children. Four kindergarten teachers completed Ready-Set for a selected child in their classrooms. Once the teachers had completed the assessment, they then completed a survey designed to evaluate selected test features. The survey addressed the scoring procedure, item content, criterion clarity, and usefulness of the tool for its intended purposes. Findings indicate that teacher participants agreed that Ready-Set goals and criteria were functional, teachable, and understandable. For details, see Stevenson (2019).

AEPS-3 CURRICULUM FIELD TEST STUDY

As noted in the Introduction, the AEPS-3 Curriculum has undergone significant changes. During 2020 and 2021, a preliminary validation study of the curriculum was completed. The purpose of the study was to have teachers/providers use the AEPS-3 Curriculum in order to evaluate its utility for planning and the usefulness of the learning activities and strategies. In addition, we sought to determine perceptions of teachers/providers regarding the appropriateness of the progress monitoring suggestions contained in the curriculum.

The curriculum validity study took place in Kentucky and Oregon. Sites included an inclusive center-based early childhood program that serves children with disabilities and their typically developing peers ages 6 weeks to 6 years, and three regional EI/ECSE programs serving children who are eligible for EI/ECSE services and their families in home and classroom settings. These programs currently use AEPS as their primary curriculum-based assessment.

Eleven early childhood educators, early interventionists, and early childhood special educators participated in the study. One participant had an associate's degree in ECE, and the other 10 participants had master's degrees in ECE, EI/ECSE, occupational therapy, or speech-language pathology. The participants had between 2 and 20 years of experience working with children. A total of 25 children were included in the study. Of these children, 13 (52%) were identified as eligible for EI/ECSE services and 12 (48%) were either typically developing or at risk for developmental delay. The children ranged in age from 1 year 2 months to 5 years 6 months with 16 males and 9 females. The majority of the children were Caucasian (n = 12), six children were biracial, four children were Asian, one child was Latino, one child was African American, and one child's ethnicity was unknown.

Before data collection, teacher/provider participants were required to complete a 1-hour AEPS-3 online training module that included a narrated slide presentation on features of AEPS-3. The module highlighted features that remain the same between the second and third editions and those that are new to AEPS-3; included the scoring rules and guidelines; and included video clips that allowed participants to practice scoring AEPS-3 Test items. Participants were also required to pass an interrater reliability test, with a score of 80% or higher, by watching video clips of children engaged in activities and scoring target items using the AEPS-3 scoring system.

After this training, participants identified three children with whom they would use the curriculum. To the extent possible, participants selected children from three categories: 1) typically developing, 2) at risk, and 3) eligible for EI/ECSE services.

Before using the AEPS-3 Curriculum, teacher/provider participants attended a 3-hour online training session. The live training included 1) an overview of the AEPS-3 linked system framework, 2) an overview of the AEPS-3 Curriculum, 3) a review of curriculum resources (AEPS-3 Curriculum levels; routines and activities; universal, focused, and targeted strategies; forms for curriculum planning and for tracking and monitoring child progress), 4) learning activities, and 5) next steps for completing the study. After the training, participants used the AEPS-3 Curriculum to implement instructional activities and strategies with each child for 8 to 10 weeks during ongoing classroom or home routines and activities. In addition, teachers/providers were required to develop embedding schedules, identify strategies to address targeted skills, and complete tracking and progress monitoring forms to submit to research staff.

After using the AEPS-3 Curriculum with their selected children, participants joined online focus groups. Focus group questions addressed using the curriculum for its three intended purposes: planning, implementation, and progress monitoring. Teachers/providers shared their experiences using the AEPS-3 Curriculum, highlighting what they found useful and offering suggestions for improvement.

Two researchers met to determine congruence on codes from the data and to identify themes. The data from the teachers/providers in the focus group was transcribed and then individually coded, identifying four themes within the data.

First, teachers/providers described their process for using the AEPS-3 Curriculum. They used the child's developmental age to identify which curriculum level to use, and they used the skills matrix to identify the routines within which to embed each child's identified goals. They also relied on children's preferences to help them select routines and activities from the curriculum. The curriculum helped them link the assessment information collected from the AEPS-3 Test to the AEPS-3 Curriculum.

Second, teachers/providers perceived that the AEPS-3 Curriculum helped them implement recommended practices for young children with and without disabilities. Specific practices they mentioned included being consistent and intentional in supporting children's development, linking assessment and instruction, embedding children's goals into ongoing routines and activities, and supporting families to implement instruction in routines and activities.

Third, teachers/providers expressed their belief that the AEPS-3 Curriculum provides support for implementing a multi-tiered system of support (MTSS) for young children with and without disabilities. Specifically, the AEPS-3 Test provides guidance about not only which skills a child needs to learn but also which tier of support the child needs to learn those skills. Teachers/providers also found that they typically provided instruction at the universal tier and then built upon that tier as is recommended when implementing MTSS.

Finally, teachers/providers found that it was feasible to implement the progress monitoring strategies associated with the AEPS-3 Curriculum within their programs.

User Feedback

AEPS users provided feedback during 2019 and 2020 on various AEPS-3 components, including the Spanish translation, Assessment Activities, Family Report, and FACS.

Spanish Translation

Bilingual professional translators who have extensive experience working with children and families completed the translation of the AEPS-3 Test items, criteria, examples, Family Report, and FACS into Spanish. Facilitating the translation was a bilingual expert who develops Spanish language and early literacy assessments and delivers training to improve assessment and intervention practices with dual language learners. More information about the translators appears on the AEPS-3 Forms USB in About the Spanish Translation.

To ensure the Spanish translation's accuracy, both translators reviewed all translations, as did a speech-language pathologist who works with Spanish-speaking families in early intervention and who has spent time administering AEPS with children in Guatemala. The translators also met with an AEPS author to review the intent of goals and objectives that needed clarification to ensure that their intent remained intact after translation. Discrepancies in the translation were then reconciled, with agreement from the translators. Items in the Social-Communication area of AEPS-3 were reviewed carefully to ensure that they were appropriate for the Spanish language and that the items captured the syntactical features of Spanish.

A focus group of six mothers and one father whose children received ECSE services provided feedback. Participants were from Mexico, Guatemala, and Puerto Rico. Families reviewed the Spanish translations of the AEPS-3 Test, Family Report, and FACS content for clarity and cultural appropriateness. In general, families reported that items were easy to understand and said they could answer all items. Families were also asked to read the forms and write notes on any items they felt needed revision. The translators then discussed the items with the group, and the facilitator took notes and documented all concerns and recommendations. The translators systematically reviewed and revised any items identified as problematic, with recommendations from the author and facilitator.

During the translation process, the translators used their extensive experience working with Spanish-speaking families of young children to focus on developing a translation that a broad range of Spanish speakers could understand, deliberately choosing vocabulary that was not overly specific to a particular region or country. To ensure that the translation is as widely accessible as possible for families, native Spanish speakers of Chilean, Costa Rican, Mexican, Puerto Rican, and Uruguayan backgrounds also reviewed the Spanish translations of the AEPS-3 Test, Family Report, and FACS for accuracy.

Assessment Activities

Pilot testers from Iowa and Kentucky used the 10 AEPS-3 Home-Based Assessment Activities and 10 Center-Based Assessment Activities with 19 children from birth to age 5 years, both typically developing and at risk for developing or with disabilities. The 3 Ready-Set Home-Based Assessment Activities and 4 Center-Based Assessment Activities were pilot tested with children ages 4–6 years. Pilot testers reported that instructions and format are clear and that all assessment activities are easy to implement and helpful. Modest revisions were made to a few assessment activities in response to tester feedback and suggestions for improvement.

Family Report and FACS

Pilot testers from Iowa and Kentucky used the Family Report with families of six children ages 1–4 years, typically developing and at risk for developing or with disabilities. Testers reported that the Family Report format is easy to follow, questions are useful for their work, the form is accessible for families and provides information that is useful for understanding family values and challenges, and information collected is useful for targeting teaching content. They noted that some parents may need support to complete the Family Report and that follow-up is needed for developing an IFSP/IEP.

Pilot testers from Iowa, Kentucky, South Carolina, and Virginia used the FACS with families of 10 children ages 1 to 4 years, typically developing and at risk for developing or with disabilities. Pilot testers found that the FACS format is easy to use, questions are generally useful, and it is written at a level that is friendly for parents. Testers provided specific feedback to make some examples clearer for a wide range of families.

INTERNATIONAL RESEARCH

AEPS-3 components have been translated into Chinese, French, and Spanish, and translations into other languages are underway. This section addresses work that has been conducted using the French translation in Canada.

French Translation Research

AEPS-3 has been translated into French and in that language is referred to as Programme EIS. Carmen Dionne and her colleagues and students at the University of Québec à Trois-Rivières have conducted a series of investigations examining features of the French translation of the AEPS-3 Test and associated components. Some studies used preliminary versions of AEPS-3. Completed studies have addressed interrater reliability and content validity of EIS/AEPS-3. In addition, a study conducted by Dubé (2019) examined the usefulness of the Child Progress Record component of the EIS/AEPS-3. In general, findings suggest that the French translation of AEPS-3 is psychometrically sound and can be recommended for use with populations of young French-speaking children. More detail on each investigation follows.

Interrater Reliability

Three studies (Bergeron, 2017; Guilbert, 2019; Lemire, 2013) examined interrater agreement of the French translation of AEPS-3, with Guilbert's focusing on the Gross Motor, Fine Motor, and Adaptive areas of the EIS/AEPS-3; Bergeron's, on the Fine Motor area; and Lemire's, on the Literacy and Math areas.

Guilbert's study (2019) evaluated the interrater reliability of the Gross Motor, Fine Motor, and Adaptive areas of the EIS/AEPS-3 using a series of videos showing seven children engaged in routine activities in their homes. Four children were typically developing and three had disabilities. The interrater reliability test consisted of viewing the video clips and scoring 93 EIS/AEPS-3 items from the three development areas (24 Gross Motor items, 42 Fine Motor items, and 27 Adaptive items). To establish the gold standard for scoring, two independent observers observed and scored all 93 test items and then reviewed them for final agreement. Subsequently, 22 observers who had successfully completed interrater reliability training on the EIS/AEPS-3 scored the 93 items independently on children. Scores between two independent raters were then compared. Results indicated high levels of agreement, with intraclass correlations (ICCs) in the Gross Motor area ranging from 0.648 to 0.905 ($M = 0.81$, $S = 0.07$) and ICCs in the Fine Motor area ranging from 0.648 to 0.905 ($M = 0.81$, $S = 0.07$). For the Adaptive area, the ICCs vary widely, ranging from 0.290 to 0.940 ($M = 0.705$, $S = 0.17$).

Bergeron (2017) evaluated interrater reliability in the Fine Motor area with two evaluators independently rating this area while watching a series of videos showing 20 children ages 3 to 6 years engaged in assessment activities in their homes. Thirteen children were typically developing, and seven had atypical development. Before scoring the video clips, evaluators were required to have achieved an overall score of 80% or higher on the interrater reliability training videos. Evaluators used the AEPS 3-point scoring system. The results show substantial agreement between the assessors for all children ($k = .747$, $p < .001$).

Lemire (2013) and her colleagues (Lemire et al., 2014) conducted studies to determine the interrater agreement between two independent evaluators who scored the EIS/AEPS-3 Literacy and Math areas. The evaluators completed the two areas on a sample of 32 children ages 4 to 6 years (22 children with typical development and 10 with atypical development). The majority of children were assessed in a school setting. Lemire developed an assessment activity designed to collect information on all Literacy and Math items and used it as a pretest for the evaluators. The degree of interrater agreement on the pretest was greater than 80%. The 32 children were evaluated subsequently using the AEPS 3-point scoring system. To calculate the degree of interrater agreement, two evaluators were paired for each of the children assessed. The results indicate high levels of agreement for the Literacy ($kr = .82$) and Math ($kr = .73$) areas.

Content Validity

Lemire (2013) and her colleagues (Lemire et al., 2014) investigated content validity for the Literacy area of the EIS/AEPS-3. Four literacy experts completed a questionnaire to examine this area's content validity. The questionnaire consisted of eight questions repeated for each item. The results indicate that these experts found items in the Literacy area to be relevant and functional. However, these experts also noted the number of Literacy items to be insufficient.

Utility

Braconnier (2020) examined the usefulness of the Literacy and Math areas of the EIS/AEPS-3 using Grisham-Brown's utility survey (2017) with 11 experienced practitioners. Feedback from this group suggested that the goals and criteria for the Literacy and Math areas are functional, teachable, and easy to understand.

Dubé (2019) examined the usefulness of the Child Progress Record component of the EIS/AEPS-3, which is designed to help practitioners monitor child progress. Dubé conducted semistructured interviews with eight practitioners working in inclusive child care settings. Feedback from interviews suggested that the practitioners found the Child Progress Record to be consistent with the practices recommended in their child care education programs. They reported that the form would be useful for planning interventions, identifying children's developmental needs, targeting intervention content for children with special needs, and communicating with families about their children's developmental needs.