Content Validity Analyses of Qualitative Feedback on the Revised Assessment, Evaluation, and Programming System for Infants and Children (AEPS) Test

Marisa Macy[1,*], Diane Bricker[2], Carmen Dionne[3], Jennifer Grisham-Brown[4], JoAnn Johnson[5], Kris Slentz[6], Misti Waddell[2], Melissa Behm[7] and Heather Shrestha[7]

[1]*Early Intervention Management & Research Group, USA*

[2]*University of Oregon, USA*

[3]*Universite du Quebec a Trois-Rivieres, Canada*

[4]*University of Kentucky, USA*

[5]*St. Cloud University, USA*

[6]*Western Washington University, USA*

[7]*Brookes Publishing Company, USA*

Abstract: Early childhood assessment practices, procedures, and tools can lay the foundation for an effective intervention program. The purpose of this article is to report the results of a content validity study conducted on a revision of Assessment, Evaluation, and Programming System for infants and children (AEPS®) Test, a widely used early childhood assessment/evaluation instrument. A panel of early childhood and early childhood special educator experts was assembled and asked to provide qualitative feedback on the content of the revised AEPS Test. Experts were asked to address five specific questions about item content, developmental sequences, and if assessment items represented quality teaching targets for young children. Qualitative results were used to modify items, developmental sequences, and area content.

Early childhood (EC) and early childhood special education (ECSE) professionals are becoming critical consumers of measures used to assess and evaluate young children [1-4]. Increasingly EC and ECSE professionals, as well as state and federal agencies, require that assessment/evaluation instruments have defensible psychometric properties addressing validity, reliability, and utility. This growing recognition of the importance of defensible psychometrics for tests may explain, at least in part, the increase in studies designed to examine properties such as validity, reliability, and utility of existing as well as new tools. For example, Macy and Bagnato [5] reported on the psychometric properties of nine early childhood assessment measures used to determine a child's eligibility for special education services. These authors noted that each measure reported some type of psychometric data. Other recent studies have also examined the psychometric properties of early childhood measures [6-8].

Determining an instrument's validity is generally considered the first step in examining its psychometric properties. Establishing validity of an instrument is important to ensure that the test results are accurate and defensible. Kane [9] describes validity as "a property of proposed interpretations" (p. 69)—that is, validity of an instrument is concerned with determining if it measures what it was designed to measure [10, 11]. In the case of assessing young children, understanding dimensions of validity is important for several reasons. First, screening and eligibility determination measures need to have adequate validity to ensure that children are properly identified and referred for services [7]. Second, test results used for formulating IFSP/IEPs and intervention content must provide an accurate reflection of the child's developmental status in order to target appropriate goals [12]. Third, to accurately monitor child progress, measures must provide valid information on child change over time [13].

An instrument has validity when it measures what it intends to measure. Different types of validity analyses include: concurrent, consequential, construct, content, criterion-related, face, internal, and predictive. All these analyses may be important to establish the overall validity of a measure; however, when a tool is revised it may be most appropriate to begin with an analysis of

*Address correspondence to this author at the Early Intervention Management & Research Group, USA; Tel: (814) 232-0342; E-mail: marisamacy@gmail.com

its content validity. Content validity is defined as "the extent to which inferences from a test's scores adequately represent the content or conceptual domain that the test is claimed to measure" [14].

The present study was designed to contribute to the growing body of important literature addressing the psychometric properties of early childhood assessment/evaluation measures. Specifically this study examined the content validity of the third edition of the Assessment, Evaluation, and Programming System for Infants and Children (AEPS) Test using an expert panel's qualitative feedback.

PREVIOUS WORK

AEPS® is a criterion referenced curriculum-based assessment/evaluation test, which is part of a larger system that includes curricular and family involvement components. The primary purpose of the test is to assist professionals and parents/caregivers in identifying children's functional skills, goals/objectives or outcomes, planning individualized interventions, and determining if those interventions positively impact children's progress toward targeted goals and objectives. The AEPS Test has multiple uses that include initial assessment, ongoing progress monitoring, eligibility determination, designing intervention and curriculum, and program accountability [6, 15-17].

The conceptual framework of AEPS relies primarily on a developmental perspective in which early forms of behavior are seen as the platform for evolving more complex behaviors. The AEPS Test is designed to yield educationally relevant, meaningful, and functional information that can be used to formulate developmentally appropriate goals/outcomes and objectives/benchmarks for children. These goals or outcomes, in turn, link directly to content and procedures for intervention and instruction offered in the curricular components of AEPS [18]. The purpose of the AEPS is to address the need for interventionists, teachers, child care providers, and specialists to have relevant and specific information on children's developmental status in order to select appropriate and useful behavioral targets to guide intervention efforts.

The AEPS Test uses a three point scoring system. A score of 0 means the child does not yet have the skill, a score of 1 means it is an emerging skill, and a score of 2 means the skill has been mastered. There are also qualitative scoring options that permit expanding information collected on each skill. For example, if the user is not able to observe the skill directly, s/he can use "Report" to obtain information from another caregiver or person familiar with the child. Criteria are provided for each item that explain in detail the dimensions of the skill that need to be met for mastery. The AEPS Test does not use age norms. A companion curriculum for the AEPS Test offers strategies to teach young children targeted skills.

The psychometric properties of the first and second editions of AEPS® have been studied since the 1980s, and AEPS is recognized as a tool with support for its validity, reliability, and utility [19, 20]. The psychometric data on the first and second editions of the AEPS Test are briefly reviewed next.

Validity

In terms of convergent validity, the results of the AEPS Test have been positively correlated with Bayley Scales of Infant and Toddler Development [21, 22] and the Revised Gesell and Amatruda Developmental and Neurological Examination [22, 23] as well as the Battelle Developmental Inventory (BDI) [12, 23] and the Vineland Adaptive Behavior Scales [23]. Gao and Grisham-Brown [24] examined the concurrent validity of two areas (i.e., Social-Communication and Cognitive) with similar developmental areas of BDI-2 and reported positives correlations. In terms of the sensitivity of the instrument, the results of Noh [25] and Hsia [26] suggest that the AEPS Test is sensitive to the differences in performance between children ages 3 through 5 years old, and between typically developing children and those with atypical development.

Reliability

The inter-rater reliability has been reported as adequate to excellent, the correlation coefficients varying from 0.60 to 0.97 [20-23, 25, 26]. In addition, studies have assessed the degree of inter-rater agreement of the AEPS using Cohen's Kappa; these vary from 0.60 to 0.94 [23, 25]. Grisham-Brown *et al*. [16] reported inter-rater agreement of 87%. Also internal consistency is good, with correlation coefficients ranging from .91 to .99 [22] and Cronbach's alphas of 0.83 to 0.99 [25].

Utility

The utility of the AEPS Test was examined by multiple research studies [20-22, 27, 28]. Some studies focused on documenting the usefulness of the AEPS Test when writing individualized education programs

(IEPs) for children [29-31] while another study compared parental satisfaction between the AEPS Test and conventional testing across settings [32]. Both parents and professionals reported a high level of satisfaction when the AEPS Test is used to assess young children.

As noted, the present analyses of the third edition of the AEPS Test were focused on generating preliminary information on its content validity. Prior to discussing the procedures and findings, the guidelines used to orchestrate the revisions are presented.

REVISION GUIDELINES

In 2010, work on a revision of the AEPS (i.e., 3rd edition) was begun. The Early Intervention Management and Research Group (EMRG) and Brookes Publishing Co., the publisher of the two previous editions, guided the revisions. EMRG is a non-profit group composed of original AEPS authors and other ECSE experts dedicated to managing the continued improvement of the AEPS. Brookes Publishing is a commercial business with a focus on products for early childhood, special education, child development, and related fields. Hereafter the EMRG participants are referred to as the EMRG team and Brookes Publishing personnel as Brookes staff.

The first step in the revision process was to identify the major changes to be undertaken in the existing test, curriculum, and family involvement materials [33, 34]. These changes were guided by ongoing feedback from AEPS® users. The present study focuses on changes in the Test.

Major changes included:

1. combining Level I (infants/toddlers) and II (preschoolers) of the AEPS® Test into one continuous scale;

2. revising Test items and associated criteria, and adding examples for existing, retained items;

3. adding two new content areas: Literacy and Math;

4. revising the scoring system to enhance its sensitivity to change;

5. revising raw scores for eligibility determination, and

6. adjusting curricular and family materials.

The second step was to create a set of guidelines designed to direct revisions of the AEPS Test. These guidelines included:

1. eliminating overlapping items once Levels I and II were combined;

2. adding items to address assessment gaps—particularly at the lower and upper developmental ranges of the AEPS Test while adding two new areas Literacy and Math;

3. modifying items/criteria to improve observational and scoring clarity and subsequent teaching sequence;

4. eliminating additive items when possible;

5. reducing the number of goals within strands in each area of the AEPS Test;

6. making developmental sequences as long as possible to encompass a broad range of skills,

7. adding notes in the Cognitive, Fine Motor, and Social-Communication areas to indicate those items moved to Literacy and Math/ Numeracy areas, and

8. expanding the scoring options for both quantitative and qualitative data on each item.

The core criteria used when developing items for AEPS 2nd edition were maintained to support the integrity of the tool. These included ensuring that items were: (a) measurable, (b) appropriate and functional teaching targets, (c) arranged in logical teaching sequence, (d) constructed to elicit generative responses, and (e) appropriate for assessing performance and monitoring progress.

PRESENT CONTENT VALIDITY ANALYSES OF QUALITATIVE FEEDBACK

First, the EMRG team completed initial revisions of the existing six areas. Work was begun on the new areas (i.e., Literacy and Math); graduate students of EMRG members were retained to create initial item pools. From 2010 to 2012, all areas were reviewed and revised numerous times. In 2012, the EMRG team and Brookes staff determined that the areas were sufficiently developed to begin examining the psychometric properties of the revised instrument.

To initiate the psychometric update of the revised AEPS Test, a content validity study was designed and

executed. As noted, the purpose of the study was to gather qualitative feedback on the content of the revised test from experts in early development. Specifically, panels of experts were asked to evaluate the revised content of the Test areas, focusing on these five considerations:

1. accuracy of developmental order of the Test items within the test structure;

2. gaps and/or redundancies in goal content within each developmental area;

3. clarity of Test items and associated criteria;

4. relation of Test items to content or skill, and

5. strength of Test items as functional and teachable skills.

METHOD

The study was divided into three phases: (1) planning, (2) selecting external reviewers, and (3) analyzing and incorporating feedback. Members of the EMRG team and designated Brookes staff participated in all phases.

Phase 1: Planning

This phase entailed designing the study to examine the content validity of the revised six areas of the AEPS Test and two new areas of Literacy and Math by gathering qualitative reviews from experts. The selected strategy was to assemble a group of EC experts, ECSE experts, and specialists (e.g., speech language pathologists, physical therapists, occupational therapists) to review the revisions and offer their qualitative feedback. Qualitative research designs offer open inquiry and interpretive research that can be used to allow for different ways of knowing and understanding [35-37]. The use of expert panels to address content validity has been used previously [38, 39]. For example, Schilling *et al*. [40] employed a model using an expert panel in order to retain, eliminate, or revise assessment items. Their findings resulted in validating an instrument developed to measure self-management. Another study used a similar methodology to investigate the use of an early literacy instrument (i.e., Dynamic Indicators of Basic Early Literacy- DIBELS) on children who were deaf [41].

The EMRG team and Brookes staff collaborated to create a pool of potential expert reviewers, and

Brookes staff sent letters of invitation to all potential reviewers with accompanying materials and specified timelines. A plan was constructed in which the EMRG team would be blind to the status of the expert panel and individual reviewers. From the reviewers who responded, a final group of experts was to be selected to participate in the study. Expert reviewers were to be selected for each developmental area of the Test, and a separate group of reviewers was to be selected to examine the entire Test. Feedback from these reviewers would be used to examine the content validity of the revised AEPS® Test and make adjustments as necessary.

Phase 2: Expert Reviewer Selection Process

To begin Phase 2, the EMRG team and Brookes staff generated a list of potential expert reviewers. Criteria for expert reviewers included: holding an advanced degree; completing relevant work in the area (i.e., published papers, and research in EC, ECSE, or a related field); and meeting project timelines. Reviewers did not have to be familiar with the AEPS. Brookes staff created the final groups of reviewers, ensuring that the EMRG team did not know who the publisher selected to participate on the expert panels. Expert reviewers (4 to 5 participants per area) were assigned one developmental area (e.g., Gross Motor, Fine Motor, and Social-Communication) matched to their area of expertise. Some experts were asked to review the entire Test. This latter group was composed of individuals with broad-based knowledge of child development and intervention.

In 2012, potential reviewers were sent an introductory request by email along with explanatory materials outlining the task, timelines, required steps to complete, description of the conceptual framework of the AEPS Test, its organization, and developmental content. The goal was to obtain a minimum of 4 reviews per developmental area and 4 reviews of the entire scale.

If a potential reviewer declined, Brookes staff contacted another reviewer for that area (e.g., Adaptive) from the master list. Additional reviewer names were identified and contacted as needed. Only one Brookes staff member knew the identification of the final set of reviewers. This person was not involved in the analyses of feedback provided by reviewers. The final group consisted of experts in early intervention/early childhood special education, child development, speech/language pathology, physical and occupational therapy.

Table 1: Areas by Number of Reviewers Contacted & Number Who Returned Reviews

Area	# Reviewers Contacted	# Received Reviews
Entire Test	13	4
Adaptive	8	4
Cognitive	8	4
Fine Motor	6	5
Gross Motor	13	4
Literacy	8	4
Math	13	5
Social-Emotional	9	4
Social-Communication	10	4
Total	88	37

Table **1** outlines the AEPS Test areas, the number of potential reviewers contacted per area (N = 88), and the number of experts who met timelines and whose qualitative feedback was included in the analysis (N = 37). All reviewers who provided feedback were included in the analyses.

After agreeing to serve as a content reviewer and submitting a signed Non-Disclosure Agreement, each reviewer received a Word document of the area(s) he/she was to review and a copy of the complete list of areas, strands, goals, and objectives for the revised AEPS Test. Brookes provided a modest honorarium to reviewers who completed the requested work.

Phase 3: Analyzing and Incorporating Feedback

The expert reviewers were asked to provide written responses to each of the five feedback questions. The questions were:

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

2.   Are there gaps and/or redundancies in goal content?

3.   Are there items and/or criteria that lack clarity?

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

5.   Are all items functional and teachable skills?

In addition, reviewers were told that they could insert comments associated with specific items directly on their copy of the area (e.g., Fine Motor). They were

asked to mark suggested changes and comments in order to retain a record of all feedback reviewers chose to provide.

The form of qualitative feedback received by Brookes varied across expert reviewers. Most reviewers answered the five posed questions separately and also wrote comments on their copy of their assigned area(s) that they had been sent. A few reviewers only addressed the five evaluation questions and did not make additional comments on their area of the test. Finally, some reviewers developed separate documents that contained their feedback. All feedback and forms were returned directly to the one Brookes staff member who knew of the reviewers' identities. Once feedback was received, the Brookes staff member removed identifying information from all documents. Then the feedback was transmitted to the EMRG team and selected Brookes staff work groups (described next).

The analysis was of a qualitative nature and required that each EMRG team member assigned to an area carefully review the feedback provided by the experts. As noted in Table **1**, feedback was obtained from four or five expert reviewer per area and five reviewers for the entire test. Eight work groups (i.e., one for each of the eight AEPS areas) were created. Each work group was composed of two to four EMRG team members and one Brookes staff member.

Each work group had a chair who was assigned the additional responsibility of ensuring that each work group member reviewed the expert feedback in a timely manner and that feedback from work group members was assembled properly. Work group members independently reviewed the feedback for

each area assigned and also reviewed the feedback from the experts who evaluated the overall Test using the procedure described next.

The work group chair first reviewed feedback from all experts (i.e., specific area and overall expert feedback) and made basic editorial changes if necessary (e.g., typos, grammar, format adjustments). Next the chair completed a Content Validity Analysis form using the feedback provided by each expert reviewer. For each reviewer's recommended change, the chair

- entered the item identification (G 1.0) and the nature of the feedback provided by the reviewer (e.g., goal is too advanced);

- noted whether she agreed with the recommendation, did not agreed, or felt the recommendations should be discussed, and

- entered all other comments.

The Content Validity Analysis form was then sent to a second work group member for review; this member also checked *yes* for agreement, *no* for disagreement, and *maybe* for discussion of reviewed comments. The form was circulated until all members of the work group had entered their feedback. Once all members completed the form, a phone conference including all members of the work group was held. The following actions were taken:

~   Changes were made if the recommendation by the expert reviewer was uniformly agreed to by all work group members;

~   When disagreements occurred between expert reviewer and work group participants, the item was noted and scheduled for a follow up discussion at a subsequent face-to-face meeting attended by the entire EMRG team and Brookes staff. At this meeting, the recommendations that were discussed, disagreements were resolved, and the final editing of areas was completed.

RESULTS

Qualitative feedback for the content validity analyses was gathered from two sources: a) reviewers' answers to the five feedback questions (listed above), and b) reviewers additional comments. The analyses of responses to the five questions are summarized first. Analyses of written comments provided by the reviewers follow.

Responses to Questions

All EMRG team members and Brookes staff reviewed the responses by expert panel members to the five feedback questions. Most reviewers answered no to the questions: *Are there items and/or criteria that lack clarity*, and *Are there items that are not related to content or skill*? Most reviewers answered yes to the questions: *Are the items in correct developmental order. Are there gaps and/or redundancies in goal content.* and *Are all items functional and teachable skills?* The reviewers commented on gaps/redundancies with the two levels of the Test being merged.

Written Feedback

Written comments made by expert reviewers either on the Test protocol or on a separate Word document provided the second data source for analyses. Table **2** contains a sample of written feedback for the Fine Motor area from an expert reviewer. This feedback was taken directly from materials returned by the reviewer.

As noted, all comments by reviewers were addressed by work group members and findings documented. Feedback varied across expert reviewers along several important dimensions including: philosophical differences; knowledge of and familiarity with AEPS; amount of feedback, type of feedback, and relevance of feedback. Each of these dimensions is discussed next.

Philosophical Differences

Two reviewers noted that their view of early development was different from the conceptual framework that has guided development of AEPS® Test items. For example, one reviewer noted that he/she puts more emphasis on using a social contextual lens to view early development and so felt the Test items did not reflect particularly well his/her orientation. Conversely, most reviewers shared a developmental perspective consistent with the framework and content of the AEPS Test. Feedback from these reviewers was generally incorporated by changing item wording, changing or adding examples, changing the sequences of objectives, or dividing content into different goals or objectives.

Knowledge of and Familiarity with AEPS® Test

Experts ranged from those who were familiar with the AEPS® Test to those who had little or no

Table 2:   Sample of Written Feedback for the Fine Motor Area

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

   I found a number of problems regarding the developmental sequence. Reach is the first skill a child obtains, this should be separated into swiping and directed reach and then research with grasp. The grasp items are out of order. The order is:

   1) Whole hand grasp at object

   2) Holds hand size (1 inch) object with thumb and first two fingers.

   3) Rakes small, pellet-size object and holds awkwardly in palm.

   4) Picks up pellet-size object with fingers and holds in finger pads

   5) Holds spoon or crayon with thumb and first two fingers.

   The release items are also out of order and a child generally release into a non-defined space before releasing into a container or a puzzle space. Stacking comes after release into a targeted space; however it would be good to have a least two levels for stacking as stacking 4-5 blocks/items is much more difficult than stacking 2.

   1) Release into space using whole hand opening.

   2) Release into a targeted space.

   3) Release of small object into small container.

   4) Stacking 2-3 blocks or objects

   5) Stacking 4 or more blocks or objects

   6) Correcting fitting puzzle piece into space (noted that this skill is under coordinated movements).

   Exploration and activation of objectives

   1) Bangs object on surface by waving arm.

   2) Transfers object hand to hand to explore.

   3) Holds object in one hand and activates using finger movements with other hand.

   4) Moves object within the hand using active finger movement.

2. Are there gaps and/or redundancies in goal content?

   Noted that visual motor is included under writing; it should also be part of the fine motor section.

   Why are there no cutting items? Using scissors and other tools (markers) are important fine motor skills. Cutting is important for learning to use two hands together and for learning to hold an object while moving it.

   Did not see use of spoon or fork; should be emphasized as a functional skill. The steps involve

   1) Can bring spoon to mouth

   2) Can scoop food and bring to mouth

   3) Eats with spoon without excessive spillage.

   4) Eats a variety of foods with a spoon.

   5) Uses both spoon and fork to eat.

   You should consider eye-hand coordination that leads to writing.

   Consider having a section of bilateral integration or bimanual coordination because often children have difficulty in this area.

3. Are there items and/or criteria that lack clarity?

   The items are clear and easily understood. If anything they are overly simplified and lack specificity.

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

   Coordinated movement seemed to cover different types of skills. It might be best to organize:

   1) Activates objects with two hands

   2) Activates objects within the hand using fingers.

   3) Aligning or stacking objects

   4) Turning objects and rotating objects.

   5) Puzzle assembly.

   6) Using tools appropriately: Scissors, markers

5. Are all items functional and teachable skills?

   The items seem to be teachable but not in the developmental sequence presented.

experience with the Test prior to this study. Many reviewers indicated their familiarity with the AEPS Test. For example, one reviewer noted, "The AEPS is a well-respected Curriculum Based Assessment, which has a long history of use in the field of Early Intervention. It has clear strengths in its use of skills that are both observable, and functional." A few reviewers indicated unfamiliarity with the AEPS Test. For example, a reviewer indicated he/she had difficulty understanding the item sequencing under goals (i.e., objectives go

from hard to easy to enhance teaching sequences) suggesting unfamiliarity with the Test despite being given material that described the organizational structure of the AEPS Test. Some feedback from another reviewer could not be used because of his/her misunderstanding about the age ranges. Feedback from these reviewers generally was considered not helpful in making improvements to the Test. Reviewers who understood the AEPS Test framework, content, and organization offered feedback of significant value that was usually incorporated to improve the content in some way.

Amount of Feedback

Again expert reviewers varied significantly in terms of how much feedback they provided. Some reviewers answered the five questions with either "*yes*" or "*no*" or with one or two sentences while others responded with paragraphs of feedback. Both types of feedback were considered; however, in general the more detailed the feedback, the more assistance it provided in considering changes.

Type of Feedback

This dimension referrs to the specificity of the provided feedback. Again reviewers ranged from offering very specific and detailed information to offering general observations. Many of the experts offered changes for specific items. Other reviewers kept their feedback more general. For example, one reviewer commented, "In general, I thought items in most strands were ordered in a developmentally appropriate way…." The range in type of feedback turned out to be useful in considering modifications to items and item sequences.

Relevance of Feedback

This final dimension refers to work group members' judgments concerning the appropriateness of the feedback. Even though reviewers had an outline of all the revised AEPS® Test areas, strands, goals, and objectives (and were offered access to other areas if they thought it might be relevant to their review of an area), some remarked that content was missing from a particular area. For example, in this revision of the AEPS Test, authors decided to move writing skills items from the Fine Motor area to the Literacy area. Consequently, if reviewers saw only the Fine Motor but not the Literacy area, their comments that writing skills content was missing from the Fine Motor area were

disregarded because that content existed in the Literacy area. The final version of the AEPS® Test will alert users when content can be found in an associated area.

DISCUSSION

Reviewing and incorporating expert feedback was an arduous undertaking. Much of the material offered was of great use in one of three ways [42]. First, when a consensus of reviewers noted a specific problem with a strand, goal, or objective, work group members made adjustments or modifications as recommended. Second, when there was a consensus across reviewers about a problem with developmental sequences, work group members made changes that addressed the reviewers, concerns. Third, if one or two reviewers noted a problem that resonated with work group members, some relevant adjustment was undertaken. In instances where comments reflected one reviewer's judgment, where comments lacked relevance, or where philosophical differences existed, recommended changes were generally not undertaken.

In general, feedback on the revised Test items tended to be positive. The changes to content and suggestions for reordering that were suggested were mostly relatively minor. As might be expected, feedback for the two new areas of Literacy and Math was more prolific and contained suggestions for more substantive changes.

Feedback on the Literacy area took additional time to examine because experts hold a range of perspectives on the developmental content and sequence for early literacy [43, 44]. Recommendations from reviewers were often in opposition to each other. The Math area also elicited considerable variation in feedback. These two areas have received the greatest attention and have undergone the most significant modification based on expert reviewer feedback.

This study has two significant limitations. First, the expert panel members were recommended by members of the EMRG team and Brookes staff; consequently, there may have been a conscious or unconscious bias toward nominating individuals with shared values. It is possible, perhaps likely, that other experts would have offered different feedback. It is possible that this feedback may have been more negative. Second, other more empirically based research designs and methodologies may have produced different results. Other analytical approaches,

like Rasch [45] or the Delphi technique, may have yielded outcomes far different from these qualitative analyses. The manner in which the feedback was collected and examined no doubt affected the outcomes.

Obtaining feedback from a wide range of perspectives informs the further development of this comprehensive assessment. The AEPS Test contains six developmental areas and two new pre-academic areas that encompass a range of items that have been vetted by a wide range of professionals covering these eight areas. This study revealed how different professionals and experts approach child development. Findings from this content validity study will be used to develop the AEPS® Test into a product that a team of professionals across disciplines with diverse backgrounds and training can use effectively.

The EMRG team and Brookes staff found the feedback provided by the expert reviewers to be extremely helpful in identifying content and developmental sequencing problems. This feedback resulted in a series of associated changes—changes that we believe have enhanced the validity of the AEPS® Test; however, subsequent investigations are needed to verify this impression.

FOOTNOTES

The Early Intervention Management and Research Group (EMRG) [46] is a non-profit, mutual benefit corporation (Bricker, 2009). It was created in 2005, "[t]o manage future developments associated with linked measurement and curriculum systems designed to enhance early childhood intervention offered to young children and their families, thus providing a public service and benefit" (EMRG Bylaws). In particular, EMRG is focused on managing and improving the *Assessment, Evaluation, and Programming System for Infants and Children* (AEPS®). For more information visit, http://AEPS®linkedsystem.com.

For more information, the reader is referred to the following white paper found at the Linked System website: http://www.AEPS®linkedsystem.com/pdf/white_paper_4_brief_history_of_emrg.pdf

REFERENCES

[1] Bagnato SJ, Goins DD, Pretti-Frontczak KL, Neisworth JT. Authentic assessment as 'best practice' for early childhood intervention: National consumer social validity research. Topics Early Child Spec Educ 2014; 34: 116-127.
http://dx.doi.org/10.1177/0271121414523652

[2] Copple C, Bredekamp S. Developmentally appropriate practice in early childhood programs: Serving children from birth through age 8. Washington, DC: National Association for the Education of Young Children 2009.

[3] Kowalski K, Brown RD, Pretti-Frontczak K. The effects of using formal assessment on preschool teachers' beliefs about the importance of various developmental skills and abilities. Contemporary Educational Psychology 2005; 30: 23-42.
http://dx.doi.org/10.1016/j.cedpsych.2004.05.001

[4] Pretti-Frontczak K, Kowalski K, Brown RD. Preschool teachers' use of assessments and curricula: A statewide examination. Exceptional Children 2002; 69: 109-123.

[5] Macy M, Bagnato S. Keeping it "R-E-A-L" with authentic assessment. National Head Start Association Dialog 2010; 13: 1-21.
http://dx.doi.org/10.1080/15240750903458105

[6] Bricker D, Clifford J, Yovanoff P, Pretti-Frontczak K, Waddell M, Allen D, Hoselton R. Eligibility determination using a curriculum-based assessment: A further examination. J Early Interv 2008; 31: 3-21.
http://dx.doi.org/10.1177/1053815108324422

[7] Bricker D, Macy M, Squires J, Marks K. Developmental screening in your community: An integrated approach for connecting children with services. Baltimore: Brookes Publishing 2013.

[8] Snyder PA, Hemmeter ML, Fox L, Bishop CC, Miller MD. Developing and gathering psychometric evidence for a fidelity instrument: The Teaching Pyramid Observation Tool - Pilot Version. J Early Interv 2013; 35(2): 150-172.
http://dx.doi.org/10.1177/1053815113516794

[9] Kane M. All validity is construct validity. Or is it? Measurement 2012; 10: 66-70.
http://dx.doi.org/10.1080/15366367.2012.681977

[10] Cronbach LJ. Test validation. In R. L. Thorndike (Ed.), Educational Measurement (2nd ed.) 1971; pp. 443-507.

[11] Newton PE, Shaw SD. Standards for talking and thinking about validity. Psychological Methods 2013; 18(3): 301-319.
http://dx.doi.org/10.1037/a0032969

[12] Hallam R, Lyons AN, Pretti-Frontczak K. Grisham-Brown J. Comparing apples and oranges: The mismeasurement of young children through the mismatch of assessment purpose and interpretation of results. Topics Early Child Spec Educ 2014; 34(2): 106-115.
http://dx.doi.org/10.1177/0271121414524283

[13] Bagnato SJ, McLean M, Macy M, Neisworth JT. Identifying instructional targets for early childhood *via* authentic assessment: Alignment of professional standards and practice-based evidence. Journal of Early Intervention 2011; 33(4): 245-253.
http://dx.doi.org/10.1177/1053815111427565

[14] Gall MD, Borg WR, Gall JP. Educational Research: An Introduction (6th ed.). White Plains, NY: Longman Publishers USA 1996.

[15] Bricker D. Assessment, Evaluation, and Programming System for Infants and Children, 2nd ed. (AEPS®). Baltimore: Brookes Publishing 2002.

[16] Grisham-Brown JL, Pretti-Fronczak K, Hallam R. Measuring child outcomes using authentic assessment practice. J Early Interv 2008; 30(4): 207-11.

[17] Slentz K, Horn K, Macy M. Assessment, Evaluation, and Programming System Overview: Tour of Multiple Uses. Annual Infant and Early Childhood Conference; Bellevue, Washington 2008.

[18] Macy M. Interactive online assessment options: A review of the AEPSi. International Journal of Early Childhood Special Education 2010; 2(3): 254-7.

[19] Bagnato SJ, Neisworth J, Pretti-Frontczak KL. LINKing authentic assessment and early childhood intervention. Baltimore: Brookes Publishing 2010.

[20] Slentz K. Evaluating the instructional needs of young children with handicaps: Psychometric adequacy of the Evaluation and Programming System-Assessment Level II. Dissertation Abstracts International 1986; 47(11): 4072A.

[21] Bailey E, Bricker D. A psychometric study of a criterion-referenced assessment instrument designed for infants and young children. J Early Interv 1986; 10(2): 124-134.
http://dx.doi.org/10.1177/105381518601000204

[22] Bricker D, Bailey E, Slentz K. Reliability, validity, and utility of the Evaluation and Programming System: For Infants and Young Children (EPS-I). J Early Interv 1990; 14(2): 147-60.
http://dx.doi.org/10.1177/105381519001400204

[23] Macy M, Bricker D, Squires J. Validity and reliability of a curriculum-based assessment approach to determine eligibility for part C services. J Early Interv 2005; 28(1): 1-16.
http://dx.doi.org/10.1177/105381510502800101

[24] Gao X. Validity of an authentic assessment in order to report young children's accountability data on early language, literacy and pre-math areas. Unpublished doctoral dissertation, University of Kentucky, Lexington, Kentucky 2008.

[25] Noh J. Examining the psychometric properties of the second edition of the Assessment, Evaluation, and Programming System for Three to Six Years: AEPS® Test 2nd Edition (3-6). Unpublished dissertation, University of Oregon, Eugene, Oregon 2005.

[26] Hsia T. Evaluating the Psychometric Properties of the Assessment, Evaluation, and Programming System for Three to Six Years: AEPS® Test. Unpublished doctoral dissertation, University of Oregon, Eugene, Oregon 1993.

[27] Bailey E, Bricker D. A psychometric study of a criterion-referenced assessment instrument designed for infants and young children. Journal of the Division of Early Childhood 1986; 10(2): 124-134.
http://dx.doi.org/10.1177/105381518601000204

[28] Bricker DD, Bailey EJ, Slentz K. Reliability, validity, and utility of the Evaluation and Programming System: For Infants and Young Children (EPS—I). Journal of Early Intervention 1990; 14(2): 147-160.
http://dx.doi.org/10.1177/105381519001400204

[29] Notari AR, Bricker DD. The utility of a curriculum-based assessment instrument in the development of individualized education plans for infants and young children. Journal of Early Intervention 1990; 14(2): 117-132.
http://dx.doi.org/10.1177/105381519001400202

[30] Notari AR, Drinkwater SG. Best practices for writing child outcomes: An evaluation of two methods. Topics in Early Childhood Special Education 1991; 11(3): 92-106.
http://dx.doi.org/10.1177/027112149101100309

[31] Pretti-Frontczak K, Bricker D. Enhancing the quality of Individualized Education Program (IEP) goals and objectives. Journal of Early Intervention 2000; 23(2): 92-105.
http://dx.doi.org/10.1177/105381510002300204

[32] Macy M, Thorndike-Christ T, Lin Y. Parental reports of perceived assessment utility: A comparison of authentic and conventional approaches. Infants & Young Children 2010; 23(4): 286-302.
http://dx.doi.org/10.1097/IYC.0b013e3181f1ec1b

[33] Bricker D, Grisham-Brown J, Dionne C, Pretti-Frontczak K, Johnson JJ, Macy M, Slentz, K, Thorndike-Christ T, Waddell M. Using the Assessment, Evaluation, and Programming System to meet federal accountability requirements. EMRG White Paper No. 6. Eugene, OR: Early Intervention Management and Research Group (EMRG) 2010.

[34] Bricker D, Pretti-Frontczak K, Grisham-Brown J, Johnson JJ, Macy M, Slentz K, Waddell M. Original purposes and expanded uses of AEPS®. EMRG White Paper No. 2. Eugene, OR: Early Intervention Management and Research Group 2008.

[35] Ferguson DL, Ferguson PM. Qualitative research in special education: Notes toward an open inquiry instead of a new orthodoxy? Journal of the Association for Persons with Severe Handicaps 2000; 25(3): 180-5.
http://dx.doi.org/10.2511/rpsd.25.3.180

[36] Ferguson PM, Ferguson DL, Taylor SJ. Interpreting Disability: A Qualitative Reader. New York: Teachers College Press 1992.

[37] Patton MQ. Qualitative Research and Evaluation Methods. Thousand Oaks, CA: Sage 2015.

[38] Dixon M, Wang S, Calvin J, Dineen B, Tomlinson E. The panel interview: A review of empirical research and guidelines for practice. Public Personnel Management 2002; 31(3): 397-428.
http://dx.doi.org/10.1177/009102600203100310

[39] Handels RH, Wolfs CG, Aalten P, Bossuyt PM, Joore MA, Leentjens AG, Severens JL, Verhey FJ. Optimizing the use of expert panel reference diagnoses in diagnostic studies of multidimensional syndromes. BMC Neurology 2014; 14.
http://dx.doi.org/10.1186/s12883-014-0190-3

[40] Schilling LS, Dixon JK, Knafl KA, Grey M, Ives B, Lynn MR. Determining content validity of a self-report instrument for adolescents using a heterogeneous expert panel. Nursing Research 2007; 56(5): 361-6.
http://dx.doi.org/10.1097/01.NNR.0000289505.30037.91

[41] Luckner JL. Using the dynamic indicators of basic early literacy skills with students who are deaf or hard of hearing: Perspectives of a panel of experts. American Annals of the Deaf 2013; 158(1): 7-19.
http://dx.doi.org/10.1353/aad.2013.0012

[42] Macy M, Grisham-Brown J, Dionne C, Johnson J, Bricker D, Slentz K, Waddell MA. Content validity study on the AEPS® third edition. Poster presented at the 7th Biannual Conference on Research Innovations in Early Intervention (CRIEI) 2014; San Diego, California.

[43] McCardle P, Chhabra V. The Voice of Evidence in Reading Research. Baltimore: Brookes Publishing 2004.

[44] National Research Council. Preventing Reading Difficulties in young Children. National Academy Press. Washington DC 1998.

[45] Winchell B. A critical examination of the technical adequacy of a curriculum-based assessment using Rasch analysis. Unpublished dissertation, Kent State University, Kent, Ohio 2011.

[46] Bricker D. (with assistance from Grisham-Brown J, Dionne C, Pretti-Frontczak K, Johnson JJ, Macy M, Slentz K, Thorndike Christ T, Waddell M.) A brief history of the Early Intervention Management and Research Group (EMRG). EMRG White Paper No. 4. Eugene, OR: Early Intervention Management and Research Group 2009.