Utility and Validity of Authentic Assessments and Conventional Tests for International Early Childhood Intervention Purposes: Evidence from U.S. National Social Validity Research

Deborah D. Lee[1,*], Stephen J. Bagnato[2] and Kristie Pretti-Frontczak[3]

[1]*The Pennsylvania State University, USA*

[2]*University of Pittsburgh/Office of Child Development, USA*

[3]*B2K Solutions, USA*

Abstract: The reported U.S. incidence of delay/disability in young children, and thus need for services, is far higher than those currently receiving early intervention supports and services [1]. Government representatives and policymakers in the U.S. have concluded that traditional tests fail to capture sufficient numbers of young children who must access early intervention supports at a critical life moment [2], even though state regulations often mandate their use. The early intervention field regards authentic assessments as a more effective alternative. However, few U.S. studies have been conducted to compare and validate the use of either conventional tests or authentic assessments for early intervention purposes. National social validity research in the United States by Bagnato *et al*. [3,4] revealed that authentic assessments fulfill the qualities/needs of the early childhood intervention field better than conventional tests. However, no national studies had been conducted to examine the qualities and patterns of use for authentic and conventional measures among interdisciplinary professionals. Based on an expanded national internet survey in the current follow-up research, we compared the qualities/patterns of use among professionals for both types of measures in the early childhood intervention field. Overall, median ratings indicate that the top authentic assessments are more valid/useful than the most popular conventional tests to accomplish most early intervention purposes. Based on the results, we share the implications as "practice-based research evidence" to guide international policymakers, professionals, and parents to advocate, choose, and use "best measures for best practices."

Keywords: Authentic assessment, early care and education, early childhood intervention, developmentally-appropriate, assessment for developmental disabilities, early childhood special education.

THE INCIDENCE OF DISABILITY AND THE NEED FOR SERVICES

As poverty has increased in the U.S., the incidence of developmental delay and disability has increased — a rate now estimated at a range of between 3 and 21% [5]. Similarly research by the WHO/UNICEF has determined that the global disability rate is 15% [6,7]. In conjunction, numerous research studies highlight the insidious negative impact of cumulative adverse childhood life events and associated "toxic stress," particularly the effects of poverty, on overall child development, school success, youth and adult physical health, and successful adaptation in life [8]. The more chronic and recurrent the adverse experiences in a child's life, the higher the risk for toxic stress and future neurodevelopmental, behavioral, learning, and chronic medical problems [9].

Despite these well-known statistics, the reported national incidence of delay/disability in young children, and thus need for services, is far higher than those currently receiving early intervention supports and services [1]. The Office of Special Education Programs reported that only 2.5% of our youngest children in the U.S. (321, 894) were declared eligible for early childhood intervention services out of a total population of nearly 13 million [2]. Current estimates of the incidence of developmental/learning/behavioral disabilities and disorders in the U.S. is 17% [1].

INADEQUACY OF TRADITIONAL EARLY DETECTION METHODS

Early interventionists, policymakers, and researchers alike, recognize that traditional methods of detecting developmental delay fail to capture sufficient numbers of young children who must gain access to critical early intervention services and supports [2]. In this recognition, the U.S. Department of Education, Office of Special Education Programs (OSEP) has funded institutes (e.g., TRACE--Tracking, Referral, and Assessment Center for Excellence) to explore the evidence-base for both conventional and alternative strategies for determining the eligibility of infants, toddlers, and preschoolers for Part C and Part B early intervention services.

The TRACE Center satellite in Pennsylvania has also produced research syntheses and selected practice guides on various methodologies for early intervention eligibility: Conventional tests and testing

*Address correspondence to this author at the Pennsylvania State University, USA; Tel: 814-863-9609; Fax: 814-863-0299; E-mail: dzl147@psu.edu

E-ISSN: 2292-2598/15

[8]; authentic assessments [8]; team assessment models [9]; assessments of social and self-regulatory deficits [10]; and presumptive eligibility [11].

Determining the true extent of delay/or functional disability for infants and young children is difficult due to many factors, but particularly, to the inadequacies of most conventional developmental measures which render them inappropriate for young children with significant functional limitations: Lack of universal design features; scripted examiner and child behaviors; unrepresentative standardization samples; and absence of prospective disability-specific field-validations [8,12-14].

OSEP indicates that "gaps in dissemination, tools, practices, training, and policy inhibit the early identification of children with special needs (p.1)" [2]. Moreover, the President's Commission on Excellence in Special Education advised that "Eligibility determination is too complicated and expensive…and efforts should be made to implement research-based, early identification and screening" (p. 24) [15].

VALID USE OF ASSESSMENT FOR EARLY CHILDHOOD INTERVENTION PURPOSES

In early childhood intervention, interdisciplinary professionals use a variety of measures to: Conduct population-based screenings for delay/disability; determine eligibility; plan and monitor individual programs and progress for young children with risks, delays, and disabilities; and document program accountability. Based on the assessments results, practical and often "high-stakes" decisions are made that profoundly impact a child's educational experiences and expectations about the child's capabilities and potential: "Misrepresenting children by mismeasuring them denies children their rights to beneficial expectations and opportunities" (p. 198) [16]. Testing or assessment results are used to link to individualized interventions for children with specific needs. As a result, the professional responsibility to select the highest quality and valid measures (e.g., the right tool for the right purpose) is not only practically important, but it is also a civil rights matter with our most vulnerable children. In order to design individual plans for care, instruction, and therapy, the overarching and primary purpose of any assessment in the field of early childhood intervention is to link assessment results to beneficial interventions [3]. By ensuring valid connections among assessment content, results, and individual programmatic goals, professionals are able to provide the vital support that is unique to each child and best protects one's rights and promotes his or her possibilities for progress.

Bagnato, Neisworth, and Pretti-Frontczak developed the following eight professional LINK standards that are reflected as the "best measures for best practices" in the field: Acceptability, authenticity, collaboration, evidence, multi-factors, sensitivity, universality, and utility [3]. The purpose of the standards is to uphold developmentally-appropriate, professionally-sanctioned, and research-based practices using authentic assessments by matching the assessment to each child. Unlike conventional tests, authentic assessments, the developmentally-appropriate alternative, capture a more accurate, holistic, and contextualized portrait of each child's profile of assets and needs [3].

Firstly, authentic assessments profile each child's functional skills which establish an individual baseline of strengths and deficits. A singular baseline for each child allows professionals to design an individualized intervention and individual goals for tracking his or performance and progress. Secondly, parents and other caregivers—most familiar and knowledgeable about the child—provide informed observations and judgments about the child's typical capabilities across natural settings and routines. Lastly, multiple samples of behaviors across time, people, and situations *via* authentic assessments ensure the most complete understanding of the child in contrast to brief, decontextualized and one-time testing. Quality standards help professionals to choose and use the best measures for each child and to make more informed and representative judgments about the competencies of each *individual* child in real-life circumstances.

FRAMING THE CURRENT RESEARCH STUDY ON ASSESSMENT IN EARLY CHILDHOOD INTER-VENTION

Previous research by the authors "set the stage" for this current expanded study on a comparison of the extent to which specific conventional and authentic measures are "socially valid" and fulfill the purposes for evaluation in early childhood intervention [3,4,17]. From previous consumer-driven, national social validity studies, Bagnato *et al*. [3,4] determined that qualities of authentic assessments are more highly regarded than conventional tests and testing procedures. Authentic assessment measures received significantly higher ratings than conventional tests from interdisciplinary professionals who work in early childhood intervention

programs. Furthermore, among the significant mean differences of authentic and conventional tests across the LINK standards, mean ratings were higher for authentic assessments. Results provided further support for the exclusive use of authentic assessments that have a valid, "practice-based evidence" profiles for intervention and research.

Despite evidence for using authentic assessments for greater validity, conventional tests are still being readily used in early childhood special education programs and services and to determine eligibility for early intervention. Macy, Bagnato, Macy, and Salaway [18] identified the following six necessary characteristics of conventional tests and testing practices to obtain accurate representations of young children's competencies and skills for early intervention eligibility: Disability sample, flexible procedures, comprehensive coverage, graduated scoring, functional content, and item density. Macy *et al*. [18] found that conventional tests lacked the six characteristics; moreover, no conventional tests had procedural flexibility. Such findings are of great concern because conventional tests are readily used to determine eligibility of a child for early intervention and early childhood special education. However, there is a lack of treatment validity studies for the most regulated tests used in early childhood intervention.

PURPOSE OF CURRENT STUDY

In this era of evidence-based practices and accountability in education, health, and human services, professionals, parents, and policymakers need to be informed about and to understand the extent to which tests can be relied upon to accomplish official purposes in the field of early childhood intervention, especially the high-stakes purpose of determining eligibility for critical early intervention services.

Expanding on the results of the previous studies cited above, our current descriptive research is a follow-up study based upon a more in-depth analysis of the qualities and patterns of use for specific authentic and conventional measures to fulfill essential early intervention purposes. Patterns of use for specific measures examine and report median ratings, frequency of usage, and exploration of various purposes, contexts, and programs. Even though research shows evidence for choosing authentic assessments over conventional tests, treatment validity studies must be conducted to examine the actual use and application of assessment results among professionals in the field. Our current research is a step in this direction.

The overarching purpose of the current study is to conduct expanded national social validity research based on the reports from "consumer-users" of specific conventional tests and authentic assessments of early development in order to appraise their capacity to fulfill the essential purposes for measurement in early childhood intervention including early detection, eligibility determination, intervention planning, progress monitoring, and program accountability required in IDEA [19].

Few nationwide studies have been conducted to validate the use of authentic assessments and conventional tests. Previous research demonstrates that insufficient studies have been conducted to establish the reliability and validity of conventional tests [18]. These studies were generally conducted with small samples for the purpose of establishing psychometric properties. Although sound psychometrics of tests are essential, the actual purpose for evaluation and testing must be considered for valid use of assessments by professionals. More specifically, test-users ought to be educated in administering, interpreting, and applying the results.

Therefore, the objective of this study is to provide social validity evidence for specific early childhood measures about their patterns of purpose-based use, using a national sample. This study explores answers to the following questions:

- Are the overall median ratings of the top five Authentic Assessments (AA; Curriculum-referenced and Curriculum-Embedded) higher or lower than Conventional Tests (CT)?
- For subtypes of assessments, are the median ratings for both curriculum-referenced curriculum-embedded top rated assessments are higher or lower than those of the top conventional tests?
- Across Early Childhood Intervention Programs and Professional Roles, how do Authentic Assessments (CT) compare to Conventional Tests (CT) across programs?
- What was the primary way professionals learned about using the AA and CT measures?

- What was the primary purpose and reason for using either an AA or CT measure?

METHOD

Sample

The data for this study were extracted from the original national social validity study and internet survey in 2008-2010 to examine whether early childhood intervention professionals were using developmentally appropriate assessments based on professional standards [3,4]. There were 1445 individual consumer social validity quality ratings collected from 969 survey respondents from 22 U.S. states in an internet-based, electronic survey housed at Kent State University (OH) facilities.

Procedures and Formats

The ratings of 80 early childhood measures were included in the survey for analysis. The sample consisted of 61 authentic assessments and 19 conventional tests. The measures were further subdivided into 3 subtypes: Curriculum-Referenced (*n*=39), Curriculum-Embedded (*n*=23), and Conventional tests (*n*=19). Measures were rated on a complex, online electronic rubric of eight operationally-defined LINK standards set on a series of 5-point rating scales ranging from 1 (*Unacceptable*) to 5 (*Exemplary*).

The total number of ratings received was 1445. However, two responses were missing data in which respondents only initiated the online survey session. To be considered a valid response, the survey entry had to include at least one complete set of ratings for one of the eight LINK standards. After deleting the two invalid responses, the total number of responses included in the analysis was 1443. Pairwise deletion was used to exclude cases with missing data in the analyses. For the purposes of this study, "incidents of judgment' were used as the units-of-analysis when examining the ratings of specific assessments so having the same respondents evaluating several instruments is not a relevant consideration.

RESULTS

Authentic Assessments (AA) and Conventional Tests (CT)

The Overall Median Ratings of the Top Five Authentic Assessments (AA) are Higher than those of the Top Five Conventional Tests (CT)

All of the top rated AA are used for typically developing, at risk, and/or children with disabilities with regard to social-emotional development (see Table **1**). Interestingly, the top rated CT are not IQ tests. Also, most of the CT can be administered, on average, in 10-20 minutes as they are mostly used for screening purposes or for specific motor/perceptual and communication competencies among a special population of children such as children with visual impairment [20-22].

Subtypes of Assessments

Overall Median Ratings for both Curriculum-Referenced Curriculum-Embedded Top Rated Assessments are Higher than those of the Top Conventional Tests

Across the eight LINK standards, curriculum-referenced and curriculum-embedded Authentic Assessments had higher median ratings than CT,

Table 1: Top Five Authentic Assessments (AA) and Conventional Tests (CT)

	Mdn	***SD***
Authentic		
The Creative Curriculum Developmental Continuum for Infants, Toddlers and Twos	5.00	0.81
The Devereux Early Childhood Assessment for Infants and Toddlers (DECA-I/T)	4.50	0.66
The Ounce Scale	4.50	1.04
Work Sampling System®	4.50	1.40
Desired Results Developmental Profile-Infant/Toddler Instrument (DRDPI-R, IT)	4.00	1.06
Conventional		
Battelle Developmental Inventory, Second Inventory (BDI-2)	3.00	0.89
Cognitive Abilities Scale-Second Edition	3.00	0.00
Developmental Assessment of Young Children (DAYC)	3.00	0.86
Early Screening Inventory-Revised™ (ESI-R™) 2008 Edition	3.00	0.81
Miller Assessment for Preschoolers	3.00	1.32

Table 2: Top Five Authentic Assessments (AA) for Curriculum-Referenced, Curriculum- Embedded, and for Conventional Tests (CT)

	Mdn	*SD*
Curriculum-Referenced		
The Devereux Early Childhood Assessment for Infants and Toddlers (DECA-I/T)	4.50	0.66
The Ounce Scale	4.50	1.04
The Work Sampling System®	4.50	1.40
Kent Inventory of Development Skills (KIDS)	4.00	1.41
Early Development Instrument	3.50	2.12
Curriculum Embedded		
The Creative Curriculum Developmental Continuum for Infants, Toddlers and Twos	5.00	0.81
Desired Results Developmental Profile-Infant/Toddler Instrument (DRDPI-R, IT)	4.00	1.06
Desired Results Developmental Profile-Preschool Instrument (DRDP-R, PS)	4.00	1.01
Learning Accomplishment Profile-Diagnostic, Third Edition (LAP-D)	3.75	0.35
New Portage Guide Birth to Six	3.50	2.12
Conventional		
Battelle Developmental Inventory, Second Inventory (BDI-2)	3.00	0.89
Cognitive Abilities Scale-Second Edition	3.00	0.00
Developmental Assessment of Young Children (DAYC)	3.00	0.86
Early Screening Inventory-Revised™ (ESI-R™) 2008 Edition	3.00	0.81
Miller Assessment for Preschoolers	3.00	1.32

except for acceptability and evidence (see Table **2**). Such a result may reflect a continuing need for increased awareness among professionals about the use of authentic assessments. Secondly, curriculum-embedded assessments have the highest median ratings for the majority of the LINK standards. This reflects that the practicality and ease of use for curriculum-embedded assessments by professionals, especially teachers in relation to the most important purpose, is to plan and evaluate children's individualized interventions. Since the items, or a child's curriculum competencies, in curriculum-embedded assessments are the assessment contents themselves, this may help minimize additional work by professionals and the teaching team when assessing and programming for a child.

Assessments Across Early Childhood Intervention Programs and Professional Roles

Standard Deviations for Authentic Assessments (CT) are Smaller than those for Conventional Tests (CT) Across Programs

This result reflects the fact that professionals are more confident and consistent in ratings across the eight LINK standards for Authentic Assessments (AA) (see Table **3** for list of top five rated assessments used in programs). Although the Ages and Stages Questionnaire (ASQ) [23] is the most frequently rated assessment across all programs, it is not one of the top five highly rated among Authentic Assessments (*Mdn*=2.50, *SD*=1.28). Thus, the ASQ seems to be the most popular measure especially for EI-Home Based and Preschool Special Education programs. Also, the ASQ was the most widely used authentic assessment and Battelle Developmental Inventory (BDI-II) [22] was the most widely used conventional test (CT) across all professional roles. Among the professional roles, the ASQ was most frequently used by individuals in itinerant teacher/consultant (40.3%) and classroom assistant (40%) roles. Among conventional tests, BDI-II was most frequently used by individuals in therapist/ specialist (56.9%) and itinerant teacher/consultant (50%) roles.

Primary way Assessments were Learned by Professionals

Workshops Received the Overall Highest Percentage of Responses among Professional Roles, Particularly by those in Researcher/Faculty Roles

More specifically, for authentic assessments (AA), the majority of respondents indicated that they learned about the assessments through a workshop (see Table

Table 3: Top Rated Measures Used in Programs

	Authentic	*Mdn*	Conventional	*Mdn*
EI-C	The Devereux Early Childhood Assessment for Infants and Toddlers (DECA-I/T) & Functional Emotional Assessment Scale	5.00	Early Screening Inventory-Revised™ (ESI-R™), 2008 Edition	3.25
CC	The Creative Curriculum Developmental Continuum for Infants, Toddlers and Twos & The Devereux Early Childhood Assessment for Infants and Toddlers (DECA-I/T)	5.00	Early Screening Inventory-Revised™ (ESI-R™), 2008 Edition	4.50
EHS	Pediatric Evaluation of Disability Inventory (PEDI™)	5.00	Battelle Developmental Inventory, Second Edition (BDI-II)	3.00
HS	The Creative Curriculum Developmental Continuum for Infants, Toddlers and Twos & Early Development Instrument	5.00	Early Screening Inventory-Revised™ (ESI-R™), 2008 Edition	4.50
EI-H	Desired Results Developmental Profile-Revised Preschool Instrument (DRDP-R, PS), Pediatric Evaluation of Disability Inventory (PEDI™)* & Work Sampling System®	5.00	Early Screening Inventory-Revised™ (ESI-R™), 2008 Edition	4.25
ECSE	New Portage Guide Birth to Six	5.00	Early Screening Inventory-Revised™ (ESI-R™), 2008 Edition	3.50
PS-Prv	High Scope Preschool Child Observation Record	4.50	Early Screening Inventory-Revised™ (ESI-R™), 2008 Edition	4.50
PS-Pub	The Ounce Scale* & Desired Results Developmental Profile-Infant/Toddler Instrument (DRDPI-R, IT)	5.00	Learning Accomplishment Profile-Diagnostic, Third Edition (LAP-D)	4.00

Note. EI-C=EI-Center-based; CC=Childcare; EHS=Early Head Start; HS=Head Start; EI-H=EI-Home-based; ECSE=Preschool Special Education; PS-Prv=Private Preschool; PS-Pub=Public Preschool.
*Highest overall mean rating.

4). However, the majority of respondents indicated that they were self-taught and informally learned about the assessments from a colleague, especially for conventional tests (CT).

Primary Purpose and Reason for Using Measures

The Top Primary Reason for Using a Respective Measure is that it is Required by Local Or State Regulations And Mandates

Among the three types of AA measures, professionals found the curriculum-referenced assessments the best for the following reasons: Offered the most useful information for programming and understanding the child's needs; best included the family or other team members; and was the most practical, clearest, and easiest to use. Interestingly, the majority of respondents indicated conventional tests (CT) were presumed to be the most valid and reliable by professionals. More specifically, 57.1% of respondents who indicated "other" as professional role and 24.1% of respondents who indicated "researcher/faculty" as their professional role indicated validity and reliability as their primary reason for using conventional tests (CT). Such a result reflects the

Table 4: Primary Way Professionals Learned to Use an AA and CT Measure

	Curriculum Referenced	Curriculum Embedded	Conventional
Informally, from a colleague (watching, talking to others)	22% Classroom assistant	15.3% Itinerant teacher/Consultant	25.7% Lead classroom teacher
In undergraduate or graduate studies	12.1% Therapist/Specialist	6.8% Researcher/Faculty	20.8% Classroom assistant
Through a workshop	40.4% Other	55.7% Researcher/Faculty	27.1% Administrator/Supervisor
Self-Taught (Read the manual)	25.5% Classroom assistant	22% Therapist/Specialist	26.4% Itinerant teacher/Consultant

Table 5: Top Purpose for Use of AA and CT Measure

	Curriculum Referenced	Curriculum Embedded	Conventional
Screening	ASQ (82.1%)	IED2 (78.8%)	BDI-II (60.5%)
Determining Eligibility for Special Education	ASQ (61.8%)	AEPS (61.1%)	BDI-II (76.4%)
Writing or updating IFSPs/IEPs	ASQSE (79.2%)	AEPS (72.2%)	BDI-II (68%)
Planning intervention/activities	ASQ (77.3%)	AEPS (76.2%)	BDI-II (65.9%)
Monitoring children's progress	ASQ (80.6%)	AEPS (73.2%)	BDI-II (67.3%)
Program evaluation	DECA (91.7%)	DRDPPI (61%)	BDI-II (47.1%)
State accountability reporting (e.g. children's performance toward state standards)	ASQ (44.4%)	DRDPPI (57.9%)	BDI-II (55.2%)
Federal accountability requirements (e.g. OSEP child outcome, Head Start Outcomes Framework)	ASQ and ASQSE tied (60.6% and 56.5%)	AEPS and CCDC35 tied (74.4% and 75%)	BDI-II (47.8%)

Note. % is the percentage of respondents would found measure useful/appropriate/meaningful for purpose.

traditional standardization-related studies completed for conventional, norm-referenced psychometric tests by publishing companies. However, a research synthesis by Macy *et al*. [18] determined that few conventional tests have been subjected to after-publication studies in real-world conditions, especially for eligibility determination; when such studies were conducted with specific child populations and under real-life agency conditions, the reliability and validity of CT did not match those results in technical manual.

With the exception of "Determining eligibility for special education," professionals found authentic assessments (AA; curriculum referenced or curriculum embedded) to be useful, appropriate, and meaningful for each purpose for the use of assessments (see Table **5**). Although 76.4% of professionals using a conventional test found the Battelle Developmental Inventory (BDI-II) to be helpful in determining the eligibility of children for special education, the majority (61.1%) of professionals also found the following authentic assessments to be appropriate and valid for the same purpose: Ages and Stages Questionnaire (ASQ) and Assessment, Evaluation, and Programming System for Infants and Children, Second Edition (AEPS®) [24]. Bricker and colleagues conducted three studies to document the superior reliability, validity, and technical adequacy of the AEPS® compared to conventional tests in both eligibility determination and individualized program planning [24-26]. Although AEPS® is the most frequently rated authentic measure, it is not the highest rated quality measure. Most frequently used for writing or updating IFSPs/IEPs in Home-based and preschool programs, professionals indicated that their primary reason for using AEPS® is because it is required. Thus, professionals' choices of using AEPS® are not directly contingent upon the quality of the measure.

DISCUSSION

Previous studies found that authentic assessments were superior to conventional tests for early childhood intervention purposes. The current national social validity study further documents the specific professional preferences for authentic assessments (AA) over conventional tests (CT) by analyzing and profiling patterns of use to fulfill early childhood intervention purposes.

In the following sections, we discuss and translate the results of this U.S. national social validity study into specific implications for both professional practice and much needed future "practice-based research" studies to inform both policy and professional behavior.

Implications for Professional Practice

Outcome Conclusion 1

Authentic assessments (AA) were rated more highly by professionals who work in in the natural setting of the classroom (e.g. teachers), whereas conventional tests (CT) were rated more highly by professionals (e.g. psychologists; therapists, faculty researchers) who work in decontextualized, non-classroom settings.

Familiar professionals who work directly with young children in the natural classroom and home settings (e.g. lead classroom teachers, classroom assistants, and itinerant teacher/consultants) rated authentic assessments higher than conventional tests. On the other hand, unfamiliar professionals who work exclusively with young children in agency, clinic, and

hospital settings (e.g. researchers/faculty, administrators/supervisors, and therapists/specialists) rated conventional tests higher than authentic assessments. These decontextualized and contrived settings included clinic rooms, agency group rooms, and hospital follow-up clinics. Among authentic assessments (AA), curriculum- referenced assessments provided the most practical and functional information about the child, and they were inclusive of parents/caregivers and team members, practical, clear, and user-friendly.

Outcome Conclusion 2

Highest rated assessments are not necessarily the most frequently used assessments.

Interdisciplinary professionals do not uniformly identify conventional tests to be valid, meaningful, and appropriate for intended purposes. For example, the Devereux Early Childhood Assessment for Infants and Toddlers (DECA-I/T) [27] was the highest rated authentic assessment. However, the Ages and Stages Questionnaire (ASQ) [23] was the most frequently used authentic assessment, especially for EI-Home Based and preschool special education programs. Despite the highest rating of DECA-I/T, the ASQ is more frequently used for the purposes of general developmental screening, and it was found to be reliable, valid, evidence-based, and sensitive [3]. The DECA-I/T [27] is best for a more specific population of children who are at-risk for social-emotional problems. In addition, 82.1% of respondents' top purpose for using the ASQ [23] was for screening. Likewise, the Early Screening Inventory-Revised™ (ESI-R™) [21] was the highest rated conventional test, but the Battelle Developmental Inventory (BDI-II) [22] was the most frequently used.

Professionals working directly with children in classrooms (e.g. lead classroom teachers, classroom assistants, itinerant teacher/consultant) need more training and support. Many of the classroom teachers and assistants were either self-taught or learned from informal interactions. Since teachers have many existing responsibilities in the classroom, additional workshop opportunity, mentoring/training, and support for assessment application may increase their likelihood of using authentic assessments for developing interventions. Additional formal training and support (e.g. workshops, professional training) may help the professionals regularly working directly with children in the classrooms may feel more comfortable in using authentic assessments over conventional tests. With more training and support, professionals working in various programs ought to utilize top rated authentic assessments reported in this study for various purposes in ECI.

Outcome Conclusion 3

Professionals working directly with young children in classrooms and homes used combinations of both the top rated authentic and conventional measures in a majority of the programs.

Top rated authentic assessments (AA) were used in child care, EI-Home Based, private preschools, and public preschool programs. Also, the top rated conventional tests were used in child care, Early Head Start, Head Start, EI-Home Based, and private preschool programs. Thus, even though the highest rated measures are not the most frequently used among professionals, they are being used in a majority of the programs.

Among professionals and across programs, the use of more highly rated assessments indicate that there is an awareness among professionals of the most valid and "best" measures. This result is positive because it shows that recent efforts and emphasis on using more valid and developmentally-appropriate measures are being translated into professional behavior in most early childhood programs—arguably related to the emphasis on application of professional standards from the National Association for the Education of Young Children (NAEYC) and the Division for Early Childhood (DEC) of the Council for Exceptional Children and related state early learning practice standards [28].

Additionally, the smaller standard deviations of authentic assessment ratings imply that professionals were more confident about authentic assessments. Although both top rated authentic and conventional measures are being used in most programs, professionals felt more positively towards authentic assessments. In practice, all programs ought to structure their classroom times so that teachers and other professionals who practice in home and classroom settings have the administrative empowerment and support to appropriately apply authentic assessments.

Outcome Conclusion 4

Professionals used those most frequently used AA and CT measures which could be applied in a shorter period of time.

By simultaneously examining the primary purpose and reason that measures were being used by

professionals, we determined that certain assessments seem to be the most popular because they do not take long to apply, score, and interpret. For example, the BDI-II and the ASQ were the most frequently used AA and CT measures for screening purposes with short application times. Taking length of test application into consideration, results are not surprising because short application periods are more favorable among professionals especially in the busy classroom settings where probes of child performance on a much more frequent basis are necessary for responsive program planning and modifications. The highest rated assessments were not the most frequently used implying that professionals may be using more user-friendly and less time-consuming assessments to accomplish their goals.

Outcome Conclusion 5

Those measures, whether AA or CT, which aligned best with the 8 LINK standards for developmentally-appropriateness practices were those that were most valued by interdisciplinary professionals.

In general, the overall higher ratings for authentic assessments (AA) indicate that professionals across programs recognize the superiority of authentic assessments in their utility and validity for fulfilling early childhood intervention purposes. However, it is very instructive to emphasize that those conventional tests (CT) which are the most frequently used and highest rated, particularly the BDI-II, are those which have qualities which align best with some of the 8 LINK standards for developmentally-appropriate assessment—functional content; developmental sequencing; graduated scoring; multi-source information; use of adaptations; and application in some natural settings. CT measures, generally, lack developmental appropriateness in the requirement for scripted administration procedures, scripted use of unfamiliar toys/objects, and need to conduct testing in contrived settings and circumstances. In addition, high use of CT is reported by professionals primarily because CT measures are often required by law in most U.S. states for eligibility determination, accountability, and program evaluation purposes. Despite the state regulatory mandates, CT measures fail to show published practice-based validity evidence for their required use in eligibility determination and accountability [18].

Implications for Future Research to Support "Best Practices" in Public Policy

This study demonstrates the need to increase awareness of how assessments are currently being used and to design, conduct, generate, and disseminate more practice-based research evidence on the "best measures for best practices." Better dissemination of practice-based research evidence will foster a more pervasive understanding and more appropriate implementation of both authentic assessments (AA) and conventional tests (CT). Moreover, such evidence would enable policymakers, administrators, and parents to make more informed decisions and to advocate for more valid choices of measures to fulfill important early intervention purposes; such evidence would reduce the urgency and occurrence of high-stakes testing in early childhood intervention. We believe that the following areas are ripe for concentrated practice-based research in real-life settings and routines as well as under the real-life challenges of meeting agency regulations:

Requiring "to-Scale" Applications of AA and CT Measures on a National and State Basis to Document their Purpose-Validity for Specific Early Intervention Purposes: Eligibility, Programming, and Accountability

It is long overdue that the field begins to conduct and disseminate the results of applied research on the utility and validity of AA and CT measures. Moreover, it is disturbing that state regulations mandate the exclusive use of specific measures to fulfill critical and often high-stakes early intervention purposes such as eligibility determination and program accountability in the absence of utility and validity studies conducted in real-world settings and under the challenges of real-world agency and community circumstances. Laboratory research in contrived settings with highly scripted procedures (high internal validity) are to be discouraged and rejected. Research under real-world settings and conditions must be required in order to produce more generalizable implications for applying the scales (high external validity). Future research demands a partnership among government, university researchers, community agencies, and publishing companies (who benefit greatly from their promoted use of these scales) to produce such research; it would be judicious for government agencies to issue a moratorium on the exclusive use of specific scales until the results of such studies are forthcoming.

Establishing the Treatment Validity of AA and CT Measures

Practitioners and researchers in the early intervention field regularly complain that the valid use

of assessment for eligibility determination overshadows the critical use of assessment for individualized curriculum planning and performance/progress monitoring—the most important function of assessment for teachers, children, and families. Moreover, complaints also abound regarding the lack of treatment validity in most CT scales; there exists an obvious disconnect in the cost-benefit ratio regarding their use, time to administer, and the lack of connection between the content of CT measures and programmatic goals or state early learning standards. In fact, most CT measures are rated low on treatment validity by professional consumers. Future research must be conducted on the treatment validity of measures required or recommended for use in early childhood intervention. Treatment validity must be a required characteristic of all measures used in the field.

Continuing Social Validity Studies of AA and CT Measures with Professionals and Including Parent Responses to Inform Test Developers about the Requirements and Preferences of their Consumers

For future studies in social validity of assessments in early childhood interventions, researchers may interview individuals in various professional roles to better understand their assessments choices and experience in using chosen measures (e.g. ease of use and implementation for planning/therapy). Also, parents or primary caretakers of assessed children may be interviewed to get their perspective on the importance of authentic measures in their children's development, therapy, and education.

Advocating for the Development of More Universally-Designed Measurement Systems and Field-Validation/Standardization Bases which Encompass all Relevant Disabilities

Only a few AA measures and no CT measures meet the requirements for universal design—inclusion of disability-sensitive features in assessment and instruction (functional versus topographical skill content or allowance for adaptations). Future research must address the critical and high-stakes issue in the fact that many children with language, behavioral, and sensory and neuromotor limitations cannot perform on the scripted tasks/items of most AA and CT measures.

Similarly, the standardization and/or field-validation samples for most conventional tests exclude children with a diversity of disabilities. Future research must rectify this glaring omission and ensure that children similar to those being assessed have a "reference sample" in the norms; without this feature, CT measures cannot be truly touted as being applicable for and "valid" for diagnosis and eligibility determination.

Examining the Social Validity and Use of Assessments for English as Second Language (ESL) and Bi-Cultural Children in the United States and Cross-Cultural Settings

In the United States alone, schools are becoming increasingly more racially and culturally diverse. Since 2005, almost fifty percent of students in the United States can be classified as "culturally different" (p. 51) [29]. With the rapidly changing demographics in the U.S., professionals and schools ought to be more aware of diverse cultural and linguistic backgrounds. Cultural values and beliefs may influence social validity and use of assessments as parents from different cultural or linguistic backgrounds may differ in their attitudes toward assessments in providing support for their children with special needs. Therefore, future studies ought to be sensitive to such changes and consumers.

Incentivizing the Design, Development, and Field-Validation of Modern Authentic Assessment Systems Using Computer-Technological Applications and Supports

Test design and development in the area of developmental assessment is so twentieth century! The age-old model of testing children at tables with soon-to-be outdated toys and objects and highly scripted procedures is good for publishing companies who thrive on frequent and evermore expensive editions of the scales, but this is very bad for our most vulnerable children. The vast use of portable computer technologies with iPads and related video and audio applications for instruction and therapy stands in sharp contrast to the horse-and-buggy model of tabletop testing. Particularly for classroom teachers or other professionals working with students in classroom settings, administration of assessments using technology can increase time efficiency and flexibility. Again, government, university, and corporate partnerships are needed to incentivize and fund the design, development, and field-validation of more contextualized and ecologically-based measurement systems which allow professionals and parents to observe, record, archive, and assess children's ongoing skill development in classrooms, homes, and community settings.

Along with such partnerships, professionals working in classrooms need to have additional training

opportunities to effectively use technology in real-time for assessment administrations. This should be the age of real-life, and real-time observational assessments of children's ongoing development and progress linked to intervention. In fact, recent developments show that this is possible and achievable [30].

Promoting more cohesive and global public policy and practices on authentic assessment through the use of the *International Classification of Functioning-Children and Youth Version (ICF-CY)* with Infants, Toddlers, and Preschoolers who are at Developmental Risk and with Developmental Disabilities

The policies and practices of early childhood intervention and assessment is an issue of international concern. All nations have a duty to care for and provide the proper services for infants and toddlers with learning disabilities and behavioral delays and their families. In addition to the U.S., European, and Asian nations are promoting the use of evidence-based research and best practices in assessment and intervention for young children with delays and disabilities.

In the European Union (EU), the European Alliance for Families (EAF) has a mission to search for the best evidence-based practices in assessment and intervention from around the EU within the special education field and make that information accessible to the general population [31]. Early intervention practices vary across the countries within the EU. Some early intervention best practices include: Publicly funded pre-primary educational opportunities which can be provided in the child's home or at outpatient/day care facilities; pre-school language support; and Sure Start. Sure Start, vaguely similar to Head Start, is a community center for children of all ages who experience learning and behavioral delays and disabilities. The center, which targets socially disadvantaged populations, specializes in early identification of special needs and provides extensive home support [31].

Much like Europe, the policies and practices of early childhood intervention and assessment in Asia vary based on country. In countries like Laos and Cambodia with higher poverty rates, there is limited access to necessary assessment and intervention services for both children and adults [32]. Early learning and intervention in Asian cultures are unique in that there is a heavy focus on individual culture because culture is so highly respected in Asian countries. Furthermore, as mentioned earlier, access to quality care relies heavily on social, political, and economic factors [33]. In Thailand, the early childhood education system strives to draw attention to the individual child's current and emerging abilities, value the child as a whole person, and involve appropriate adults in the process of defining the child's achievement. Despite common national goals, each individual early intervention center has its own goals for the children they serve.

In spite of international policies on programmatic practices in early childhood intervention, few regulations are common for early childhood assessment; cohesive international policies and professional practices can be promoted by emphasizing the advantages of using authentic and functional assessment procedures linked to the ICF-CY to fulfill early childhood intervention purposes.

In collaboration with the WHO global initiatives on disability [6,7], Simeonsson and colleagues [34,35] have written extensively about the use of the ICF-CY and accompanying "developmental code sets" to transform and unify professional practices and policies for the assessment of children with developmental disabilities [36]. The ICF-CY is the culmination of over 20 years of international collaborative work to design, develop, and field-validate a functional classification system for use with children including infants, toddlers, and preschoolers at-risk and with disabilities. The ICF-CY provides the rationale, content, and structure to unify interdisciplinary assessment practices for young children through the use of a framework which is authentic, functional, universal, and intervention-based.

Moreover, researchers [36,37] have developed and piloted a mapping framework which generates a "crosswalk" among the item content codes of the ICF-CY and the functional competencies of evidence-based, authentic assessment measures [37]. This mapping enables practitioners and researchers to more efficiently assess the capabilities and needs of young children, to plan beneficial and ecological interventions, and to engage parents in the assessment process to ensure the use of "best practices".

The functional classification system within the ICF-CY has been field-validated and mandated through government policy for use in several countries, most notably for special education in Portugal [38] and for medicine and all human services in Taiwan [39]. International initiatives championed by WHO such as the ICF-CY and the child disability assessment initiative

must be embraced by interdisciplinary professionals in order to realize the advantages and benefits of authentic assessments for fulfilling the linked purposes of screening, eligibility determination, intervention planning and progress/program evaluation for our most vulnerable children.

Conducting Future Research and International Field Validations with Assessments in Early Childhood Intervention Programs

Because the majority of the assessments used in early childhood intervention programs have been constructed in the U.S., these measures have also been normed and standardized based on western cultural and societal norms. Future research and assessment validation studies ought to consider culturally diverse populations with the measures in order for these measures to be valid and used appropriately within each society for childhood intervention purposes.

REFERENCES

[1] Center for Disease Control and Prevention. Improving the Health of People with Disabilities. Atlanta: National Center on Birth Defects and Developmental Disabilities 2014.

[2] U.S. Department of Education. To assure the free appropriate publication education of all children with disabilities: Individuals with disabilities education act, section 618 report. Washington: Office of Special Education Programs (OSEP) 2001.

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

[4] 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

[5] Fujiura G, Yamaki K. Trends in demography of childhood poverty and disability. Except Child 2000; 66(2): 187-199.

[6] World Health Organization and UNICEF. Early childhood development and disability: A discussion paper. Geneva: World Health Organization 2012.

[7] World Health Organization and UNICEF. Developmental difficulties in early childhood: Prevention, early identification, assessment and intervention in low- and middle income countries. Geneva: World Health Organization 2012.

[8] Felitti V, Anda R, Nordenberg D, Williamson D, Spitz A, Edwards V, *et al.* Relationship of Childhood Abuse and Household Dysfunction to Many of the Leading Causes of Death in Adults. Am J Prev Med 1998; 14(4): 245-258. http://dx.doi.org/10.1016/S0749-3797(98)00017-8

[9] Bagnato S, McKeating-Esterle E, Bortolamasi P. Evidence-base for team assessment practices in early intervention. Pittsburgh: TRACE Center for Excellence in Early Childhood Assessment, Early Childhood Partnerships, Children's Hospital/University of Pittsburgh; US Department of Education, Office of Special Education Programs, and Orelena Hawks Puckett Institute 2007.

[10] Yeh Ho H, Bagnato S. Research foundations for the use of social-emotional indicators to determine access to early intervention services. Washington: USDOE, OSEP, TRACE Center for Excellence 2007.

[11] Fevola A, Bagnato S, Matesa M, Lehman C. Research review for characteristics of presumptive eligibility promoting early intervention access. Pittsburgh: TRACE Center for Excellence 2006.

[12] Benn R. Conceptualizing eligibility for early intervention services. In: Bryant D, Graham M, ed. by. Implementing early intervention. 1st ed. New York: Guilford Press 1994; pp. 18-45.

[13] Shonkoff J, Meisels S. Defining Eligibility for Services Under PL 99-457. J Early Interv 1991; 15(1): 21-25. http://dx.doi.org/10.1177/105381519101500104

[14] Bagnato S. Authentic assessment for early childhood intervention. New York: Guilford Press 2007.

[15] President's Commission on Excellence in Special Education (PCESE). A new era: Revitalizing special education for children and their families [Internet]. Jessup: Education Publications Center, US Department of Education; 2002. Available from: http://ectacenter.org/~pdfs/calls/2010/earlypartc/revitalizing_special_education.pdf

[16] Neisworth J, Bagnato S. The MisMeasure of Young Children. Infants Young Child 2004; 17(3): 198-212. http://dx.doi.org/10.1097/00001163-200407000-00002

[17] Macy M, Bagnato S, Macy R, Salaway J. Conventional Tests and Testing for Early Intervention Eligibility. Infants Young Child 2015; 28(2): 182-204. http://dx.doi.org/10.1097/IYC.0000000000000032

[18] U.S. Department of Education. Part C of the Individuals with Disabilities Education Act: Final Regulations. Pittsburgh: Federal Register 2011.

[19] Allan J. Selected Assessments That Are Used For Young Children Who Are Visually Impaired [Internet]. Tsbvi.edu. 2015 [cited 9 November 2015]. Available from: http://www.tsbvi.edu/instructional-resources/3294-selected-assessments-that-are-used-for-young-children-who-are-visually-impaired/

[20] Meisels S, Marsden DB, Wiske MS, Henderson LW, 2014. Early Screening Inventory-Revised, 2008 Edition. San Antonio, TX: Pearson.

[21] Newborg J. Battelle Developmental Inventory: Second Edition. (BDI-2). Itasca, IL: Riverside 2004.

[22] Bricker D, Squires J. Ages & Stages Questionnaires: A Parent-Competed, Child-Monitoring System (2nd ed.). Baltimore, MD: Brookes Publishing Co., Inc. 2000.

[23] Bricker DD, Pretti-Frontczak KL. The assessment, evaluation, and programming system for infants and young children: Vol. 3 AEPS measurement for three to six years. Baltimore, MD: Paul H. Brookes Publishing Co. 1996.

[24] Bricker D, Yovanoff P, Capt B, Allen D. Use of a curriculum-based Measure to Corroborate Eligibility Decisions. J Early Interv 2003; 26(1): 20-30. http://dx.doi.org/10.1177/105381510302600102

[25] Bricker D, Clifford J, Yovanoff P, Pretti-Frontczak K, Waddell M, Allen D, *et al.* Eligibility Determination Using a Curriculum-Based Assessment: A Further Examination. J Early Interv 2008; 31(1): 3-21. http://dx.doi.org/10.1177/1053815108324422

[26] California Department of Education, Child Development Division. Desired Results Developmental Profile: Infant and Toddler Instrument. Sacramento, CA: California Department of Education 2006.

[27] Division for Early Childhood. Promoting positive outcomes for children with disabilities: Recommendations for curriculum, assessment, and program evaluation. Missoula: National Association for the Education of Young Children 2007.

[28] Ford D. Culturally Responsive Classrooms: Affirming Culturally Different Gifted Students. Gifted Child Today. 2010; 33(1): 50.

[29] Bracken B. Riverside Early Assessment of Learning (REAL). MN: Houghton-Mifflin-Harcourt-Riverside Publishing Co. 2014.

[30] European Commission. Support for children with special educational needs (SEN). Brussels: Employment, Social Affairs & Inclusion 2013.

[31] Kamerman S. Early childhood care and education and other family policies and programs in south-east Asia. Paris: UNESCO 2002.

[32] Kim G, Umayahara M. Early Childhood Care and Education: Building the Foundation for Lifelong Learning and the Future of the Nations of Asia and the Pacific. ICEP 2010; 4(2): 1-13. http://dx.doi.org/10.1007/2288-6729-4-2-1

[33] Simeonsson R. ICF-CY: A Universal Tool for Documentation of Disability. J Policy Pract Intell Disabil 2009; 6(2): 70-72. http://dx.doi.org/10.1111/j.1741-1130.2009.00215.x

[34] Simeonsson R, Leonardi M, Bjorck-Akesson E, Hollenweger J, Lollar D, Martinuzzi A, *et al*. ICF-CY: A universal tool for practice, policy, and research. (Document P107). Washington: World Health Organization (WHO) 2015.

[35] Pan Y, Hwang A, Simeonsson R, Lu L, Liao H. ICF-CY code set for infants with early delay and disabilities (EDD Code Set) for interdisciplinary assessment: a global experts survey. Disabil Rehabil 2015; 37(12): 1044-1054. http://dx.doi.org/10.3109/09638288.2014.952454

[36] Kronk R. Mapping ICF-CY content to the VABS and ABAS-II [MS]. University of Pittsburgh 2004.

[37] Kronk R. ICF-CY profiles for young children with specific disabilities: Mapping item content. Conference Proceedings of the International Society for Early Intervention (ISEI). New York: International Society for Early Intervention 2011.

[38] de Miranda-Correia L. Special education in Portugal: the new law and the ICF-CY. Procedia - Social and Behavioral Sciences 2010; 9: 1062-1068. http://dx.doi.org/10.1016/j.sbspro.2010.12.286

[39] Hwang A, Liao H, Chen P, Hsieh W, Simeonsson R, Weng L, *et al*. Applying the ICF-CY framework to examine biological and environmental factors in early childhood development. Taiwan Yi Xue Hui Za Zhi 2014; 113(5): 303-312. http://dx.doi.org/10.1016/j.jfma.2011.10.004

Received on 01-10-2015 Accepted on 06-10-2015 Published on 07-01-2016

DOI: http://dx.doi.org/10.6000/2292-2598.2015.03.04.2