**Reliability Measures in Early Childhood Assessment**

This paper is designed to provide best practice guidance to teachers, providers, and administrators in assessing young children in Kentucky. A definition of quality assessment is provided, as well as several areas where errors can occur in the assessment process. Suggestions for prevention or amelioration of those errors to increase reliability of assessment data are also included. Additional resources are included in the reference list, as well contact information for the KY Department of Education (KDE), Cabinet for Health and Family Services First Steps program, the University of Kentucky College of Education, and the KY Early Childhood Data Systems (KEDS) project.

What defines a quality assessment?

Quality assessment serves two main purposes: to inform instruction and to measure student progress. Informing instruction provides a mechanism to improve the quality of instruction, programs, and services and is part of a continuous assessment model. Information gained through the assessment process guides instructional decisions and delivery. Information from assessment also is used to measure student progress for state and national evaluation of program effectiveness (Rous, McCormick, Gooden, & Townley, 2007). It is for this purpose that the KEDS project was designed. Quality assessments are the keystone to effective practice, which ultimately lead to improved outcomes and school readiness for young children. This continuum is illustrated in Figure 1 below.

Figure 1. Kentucky's Early Childhood School Readiness Support System
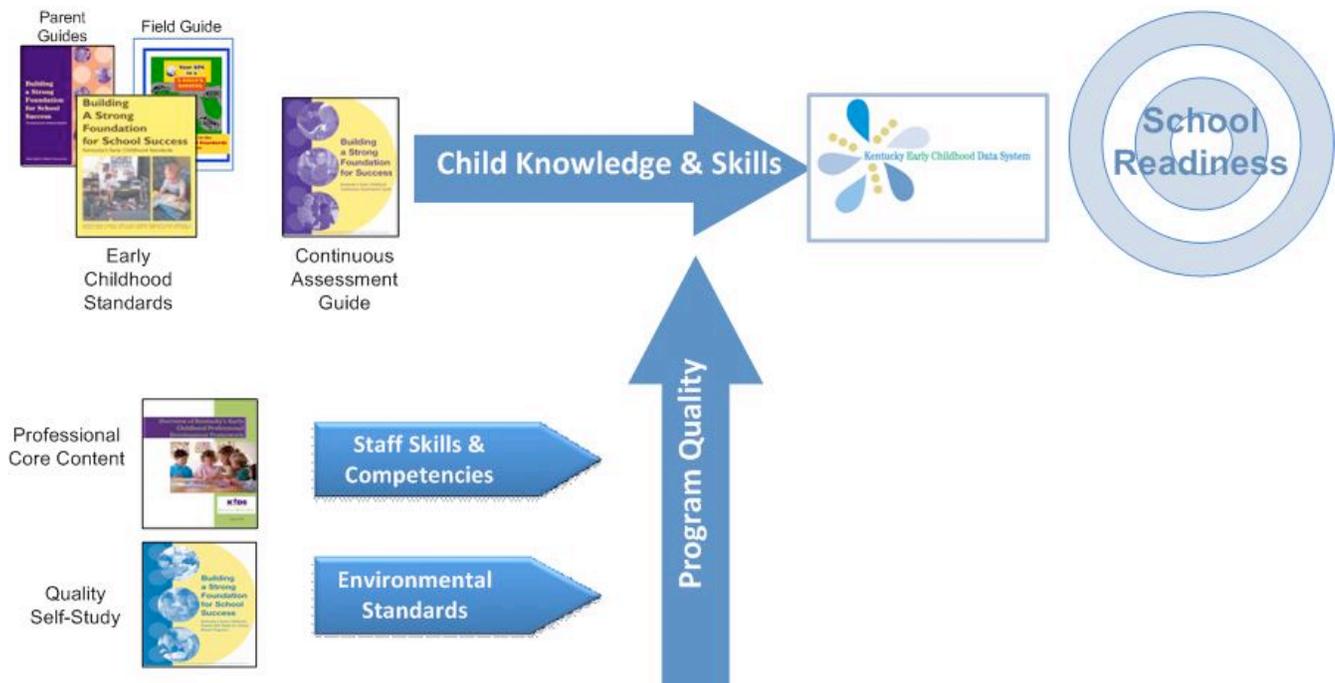
Quality assessments are described in *Kentucky's Early Childhood Continuous Assessment Guide* (2004) and include the following components:
- address relevant domains in an authentic manner;
- are ongoing and related to curriculum (Bredekamp & Rosegrant, 1995);
- are useful for planning (Bagnato, Neisworth, and Munson,1989; Bredekamp & Rosegrant, 1995);
- indicate student skills and processes needed;
- involve multiple sources of information (McCormick & Nellis, 2003);
- are understandable; maximizing learning (NAEYC & NAECDS/SDE, 2003);
- identify children for additional evaluation; and
- benefit children (Kentucky Department of Education, 2004).

What instruments best inform quality assessments?

In order to reliably measure child progress to provide useful evaluation data, Kentucky stakeholders selected a group of curriculum-based or criterion-referenced assessments (CBAs/CRAs), listed in the *KY EC Continuous Assessment Guide* in the section entitled Classroom/Instructional. These 12 tools had previously undergone an extensive approval process and were selected by a group of teachers, early intervention and technical assistance providers, state and local administrators, and university researchers. The approved instruments were chosen to provide the most useful information to guide quality programming for young children; the tools and some features are illustrated in Table 1.

Table 1: KY Approved Assessment Instruments

ASSESSMENT INSTRUMENTS			
COR-IT (6 wks-3 yrs)	**COR** (2 ½ - 6 yrs)	**Creative Curriculum** (3-5 yrs)	**OUNCE** (Birth-42 mos)
Sense of self	Sense of Self/Initiative	Social & Emotional	Social & Emotional
Social relations	Social relations	Physical	Language
Creative representation	Creative representation	Language	Cognitive
Movement	Music & Movement	Cognitive	Motor
Communication & Language	Communication & Language/Language & literacy	**50 objectives**	
Exploration & Early Logic	Exploration & Early Logic/Logic & mathematics	**with 3 steps and 3 forerunners each;**	**Six principal areas;**
28 COR items **with 5 levels each**	**32 COR items** **with 5 levels each**	**Creative Curriculum for Infants, Toddlers, and Twos** (0-3 yrs)	**2-3 questions per area;** **2 levels per question;** **for each age interval**
		Same domains	
		21 objectives; 5 steps each	
WSS (3 yrs-5th grade)	**WSS for Head Start** (3 and 4 year olds)	**E-LAP** (Birth-3 yrs)	**LAP-3** (36-72 mos)
Personal & Social Development	Social & Emotional Development	Gross motor	Gross motor
Language & Literacy	Approaches to Learning	Fine motor	Fine motor
Mathematical Thinking	Language Development	Cognitive	Pre-writing
Scientific Thinking	Literacy	Language	Cognitive
Social Studies	Mathematics	Self-help	Language
The Arts	Science	Social-emotional	Self-Help
Physical Development & Health	Creative Arts		Personal/Social
55 Indicators for 4's	Physical Health & Development		
with 3 levels each;	**55 Indicators for 4's**	**264 Items**	**348 Items**
49 Indicators for 3's,	**with 3 levels each;**		
with 3 levels each;	**51 Indicators for 3's**		
	with 3 levels each		

ASSESSMENT INSTRUMENTS			
TPBA (Infancy - 6 yrs) Cognitive Social-Emotional Communication and Language Sensorimotor **Up to 80 items for observation per domain**	**HELP** (Birth -3 yrs; 3 - 6 yrs) Cognitive Language Gross Motor Fine Motor Social Self-Help **685 developmental skills and behaviors**	**Carolina Curriculum for Infants and Toddlers with Special Needs** (Birth-3 yrs) Cognition Communication Social Adaptation Fine motor Gross motor **24 curriculum sequences**	**Carolina Curriculum for Preschoolers with Special Needs** (24-60 mos) Cognition Communication Social Adaptation Fine motor Gross motor **22 curriculum sequences**
AEPS (Birth - 6 yrs) Fine motor Gross motor Cognitive Adaptive Social-Communication Social **Birth-3 yrs: 23 Strands, 249 items** **3-6 yrs: 21 Strands, 217 items**			
Brigance IED-II (Birth-6 yrs) Pre-ambulatory Motor, Gross Motor, Fine Motor, Self-Help, Speech and Language, General Knowledge and Comprehension, Social and Emotional Development, Readiness, Basic Reading Skills, Manuscript Writing, Basic Math **1,024 items in 89 sub-domains**			

Adapted from Schroeder, C. & McCormick, K. (2005). *KY Early Childhood Continuous Assessment Guide*: Classroom/Instructional Assessment Training.

CBAs were selected to measure child progress rather than norm-referenced tools, since CBAs are more closely aligned to the *Kentucky's Early Childhood Standards* (Kentucky Department of Education, 2003)*,* and norm-referenced tools are not sufficiently congruent with these standards during the early childhood years for high-stakes decision-making (McCormick & Nellis, 2003; Shepard, Kagan, & Wurtz, 1998). The use of CBAs also allows stakeholders (including teachers, early intervention providers, and family members) to address functional areas of development for improvement for children and programs. The approved CBAs are required by Cabinet for Health and Family Services and the KY Department of Education for the measurement of early intervention and preschool student progress for reporting to the Office of Special Education Programs (OSEP).

What kinds of error can occur with assessment?

Quality assessment assumes that the instruments selected measure the skills and behaviors which are taught and are accurate; that the assessors are skilled and knowledgeable of the test and its administration; and that the results are accurately communicated. Error in assessment can be defined as a lack of generalizability due to a failure to obtain a representative sample of behaviors (Salvia & Ysseldyke, 1995). Sampling error is usually associated with norm-referenced tools, but is also helpful to consider in relation to the selection of KY-approved tools. CBAs were chosen to guide programming, as the most effective way to create an authentic link between the outcomes that are valued for children, as described in KY's standards, and the ability of assessment instruments to measure young children's skills in meeting these outcomes. At the local or district level, program directors or district personnel may also want to consider that some instruments include more items that correspond to KY standards than others; refer to the Birth to Three and Three to Four Standards Matrices in the *KY EC Continuous Assessment Guide* (Kentucky Department of Education, 2004) for a comparison of the item pool and relationship to the benchmarks for each tool.

What is instrument error?

Instrument error includes any limitations in assessment due to the nature of the particular instrument (McCormick & Nellis, 2003). For example, some of the approved instruments rely heavily on verbal responses, while others primarily allow for responses by pointing; some have great detail for scoring criteria, while others have less specificity. It is important that districts and programs select assessment tools that are best suited to their population, priorities, and skills of their staff for administration. If programs modify assessments for their specific needs, it is critical that accurate assessment results be sent to the state via KEDS in publisher-approved versions ONLY. As instruments are updated and improved, assessment data must be sent according to the most current publisher specifications (for example, with newly modified items or scoring criteria). Staff must also receive updates from certified trainers anytime instruments are modified, to ensure accurate and reliable administration.

What is administration error?

Administration error includes mistakes that persons administering assessments make due to inadequate training, poorly developed rapport with the child, or insufficient knowledge of the instrument (Salvia & Ysseldyke, 1988). These errors may include rushed administration, inadequate preparation, insufficient materials for administration, and reduced child performance due to a lack of rapport. All teachers/providers must receive adequate training from certified trainers for each instrument. In addition, when multiple teachers/providers assess children, consistent procedures for administration are critical for accurate results. Methods for collecting data and for documenting observed behaviors must be specified. Local, district or program guidelines also need to specify when assessments are to be conducted, at the least in fall and spring. KY recommended timelines for KDE districts are for fall assessments to be completed no later than October 1, and for spring assessments to be completed in April. First Steps assessments are to be completed prior to the initial IFSP, annually, and within 120 days of exit. In addition, staff must be trained in the recommended procedures to record behaviors for each instrument (i.e., by direct test, observation, interview). Programs may vary in the ways staff record such behaviors, as long as they meet publisher specifications.

What is scoring error?

As with all of the types of error described above, some errors in scoring may occur (McCormick & Nellis, 2003), including assigning of levels of mastery, consistency in scoring or crediting, tallying of responses, and simple copying mistakes. Error must be anticipated both in item crediting and in data entry for each item, with procedures put in place to check accuracy. Consistency of scoring is essential across programs and districts, and must be addressed in training and monitoring. Staff can be encouraged to check their own work, and the work of their peers. For tools that have basal and ceiling rules, these must be followed precisely. When no basal or ceiling rules exist, all items must be scored. Incomplete assessments are another type of scoring error. When staff omit scores for any item, that assessment is considered incomplete for KEDS and cannot be used for data analysis purposes. Not applicable is acceptable as a score when included as a scoring option by the publisher. Administrator/assessor bias is a factor against which to guard; teachers and providers must be trained to assess all children using the same (or standardized) criteria and can not make allowances for any children beyond what the instrument specifies. For a child's first assessment, when baseline data is being gathered, it is important for staff to complete the full assessment as early in the year as possible, so that maximal progress can be shown at the next assessment time. Assessment trainings for all tools emphasize the need to score conservatively; full credit must only be given when a teacher/provider has adequate documentation that the skills/behavior measured by the item is fully met. Tools vary on the criteria, number of observations, and/or trials that constitute mastered performance; administration manuals must be readily available and consulted for every item's criterion statement.

What are some procedures we can implement to increase reliability in assessment?

- Conduct annual trainings by certified trainers on assessment tools for new staff
- Conduct annual update trainings for returning staff, including revisions to existing tools
- Conduct and document reliability checks of each staff annually
- Provide annual training in curriculum used by each program
- Conduct trainings in anecdotal record-keeping/observations for tools using this modality
- Conduct trainings in portfolio development for tools using this modality
- Conduct trainings in online systems for staff using such systems, with computer training as needed
- Provide frequent opportunities to practice item scoring/crediting with all staff
- Provide manuals with specific scoring criteria for all staff who assess

What information will the KEDS system provide to my district or school?

KEDS is a data collection and analysis system designed to gather assessment data from a variety of programs, including preschool, early intervention, and child care agencies. KEDS does not provide child-specific data; rather, it aggregates assessment data at the program and agency level, and provides results to the responsible state agency (KDE, First Steps, and Division of Child Care). For individual child or classroom level information, programs are encouraged to use the wealth of information provided by the completion of each child's assessment. In addition, some instruments provide additional tools for evaluating classroom or program effectiveness. Online systems are especially useful in providing this level of information. Unlike current online systems which are based on generally accepted or author-selected developmental milestones, the KEDS system provides a measure of student progress toward the acquisition of KY standards at district and program levels. This information will be available for all participating programs in the 2010-2011 school year.

Cite as: McCormick, K. & Gooden, C. (2009). *Reliability Measures in Early Childhood Assessment.* Kentucky Early Childhood Data Systems, Human Development Project, University of Kentucky.

References:

Bagnato, S., Neisworth, J., & Munson, S. (1989). *Linking developmental assessment and early intervention: curriculum-based prescriptions.* Rockville, MD: Brookes Publishing.

Bailey, D.B., & Wolery, M. (2003). *Assessing Infants and Preschoolers with Handicaps.* Columbus, OH: Merrill Publishing.

Bredekamp, S. & Rosegrant, T. (Eds.) (1995). *Reaching potentials: Transforming early childhood curriculum and assessment, volume 2*. Washington, D.C.: National Association for the Education of Young Children.

Bricker, D., Capt, B., & Pretti-Frontczak, K. (2002). *Assessment, Evaluation, and Programming System for Infants and Children, Second Edition.* Baltimore: Brookes Publishing Co.

Cabinet for Health and Family Services http://chfs.ky.gov/dph/firststeps.htm

Dodge, D.T., Colker, L.J., & Heroman, C. (2001). *The Creative Curriculum Developmental Continuum Assessment System.* Washington, D.C.: Teaching Strategies, Inc.

High/Scope Educational Research Foundation. (2003). *Preschool Child Observation Record* (2nd ed.). Ypsilanti, MI: High/Scope Educational Research Foundation.

Johnson-Martin, N.M., Attermeier, S.M., & Hacker, B.J. (2004). *The Carolina Curriculum for Infants & Toddlers with Special Needs* (3rd ed.). Baltimore: Brookes Publishing Co.

Kentucky Department of Education http://www.education.ky.gov/KDE/

Kentucky Department of Education. (2003). *Building a strong foundation for school success: Kentucky's early childhood standards.* Frankfort, KY: Author.

Kentucky Department of Education. (2004). Building a strong foundation for school success: *Kentucky's early childhood continuous assessment guide.* Frankfort, KY: Author.

McCormick, K., & Nellis, L. (2003). Assessing cognitive development. In M. McLean, M. Wolery, & D.B. Bailey (Eds.) *Assessing infants and preschoolers with special needs* (3rd ed.). Saddle River, NJ: Pearson.

NAEYC and NAECS/SDE (2003). Joint Position Statement: *Early childhood curriculum, assessment, and program evaluation building an effective, accountable system in programs for children birth through age 8*. Retrieved January 27, 2004 from: http://www.naeyc.org/resources/position_statements/CAPEexpand.pdf

Parks, S. (2006). *Inside HELP: Administration and Reference Manual for HELP*. Palo Alto, CA: VORT Corp.

Rous, B., McCormick, K., Gooden, C., and Townley, K.F. (2007). Kentucky's early childhood continuous assessment and accountability system: Local decisions and state supports. In *Topics in Early Childhood Special Education.* Spring 2007, 27:1.

Salvia, J., & Ysseldyke, J. (1988). *Assessment in special and remedial education.* Boston: Houghton Mifflin.

Salvia, J., & Ysseldyke, J. (1995). *Assessment* (6th ed.). Boston: Houghton Mifflin.
Sanford, A.R., Zelman, J.G., Hardin, B.J., Peisner-Feinberg, E. (2004). *The early learning accomplishment profile* (3rd ed.), Chapel Hill, N.C.: Chapel Hill Training Outreach Project, Inc.

Shepard, L., Kagan, S., & Wurtz, E. (1998). *Principles and recommendations for early childhood assessments.* Washington, D.C.: National Educational Goals Panel. Available at http://www.negp.gov.

University of KY College of Education http://www.uky.edu/Education/

University of KY KEDS Project https://www.kedsonline.org/