EARLY CHILDHOOD
COMPREHENSIVE ASSESSMENT SYSTEM

A Partnership to Promote School Readiness by the Maryland and Ohio Departments of Education

---

A Conceptual Design of the
Early Childhood Comprehensive Assessment System
(EC-CAS)

Presentation
Rolf Grafwallner, Ph.D.
Assistant State Superintendent
Maryland State Department of Education
Stephanie Siddens, Ph.D.
Director of Early Learning
Ohio Department of Education

Association of Public Policy and Analysis and Management,
November 7-9, 2013
Washington, D.C.

Background

On December 16, 2011, Maryland and Ohio were awarded **Race To The Top-Early Learning Challenge Grants** (RTT-ELCG) in the amount of $50.0 and $69.9 million, respectively, over four years. Among other projects, these funds will support an innovative partnership to revise and enhance Maryland's and Ohio's kindergarten entry assessment, develop prekindergarten and kindergarten formative assessments, and conduct a rigorous review of existing screening tools for young children. These efforts culminated in a new comprehensive early childhood assessment system, supported by statewide technology infrastructures, and a professional development system.

A number of partners play a vital role in executing Maryland and Ohio's shared vision for improving kindergarten readiness and early childhood assessments. The main partners are Johns Hopkins University - Center for Technology in Education (CTE), WestEd, State Advisory Boards in each state, and a national Technical Advisory Committee, advising both states and facilitated by the Council of Chief State School Officers. The new assessment system will be fully implemented by both states in the fall of 2014.

Theoretical Foundations

What Does It Mean to Be Ready for School?
The importance of early childhood experiences has long been recognized as a cornerstone in the foundation for language development, cognitive development, and social and emotional development (Snow & Van Hemel, 2008; NGA, 2005; Rhode Island KIDS COUNT, 2005). Longitudinal research focusing on achievement gaps among middle and high school students has found those gaps to have been observed as early as kindergarten (Stedron & Berger, 2010). For the majority of adults, success in school (all realms, not just academic subjects) is reflected in achievements and functional abilities later in life. Monitoring children's early developmental progress ensures that the unique needs of each child are understood and provides parents and teachers with data to support development of curriculum and instructional activities that foster readiness for kindergarten and beyond.
Nearly every state in the nation has adopted policies intended to ensure that all children have access to opportunities to grow and develop from birth through age six and to start kindergarten with the academic, physical, and emotional foundation they need to succeed (Bush, 2010; Daily, Burkhauser, & Halle, 2010). But because the structure of the U.S. educational system allows for a great deal of local determination of policies and curricula, policies and practices vary widely. As of 2010, 43 states require school districts to offer a

kindergarten program, yet only 16 of those states mandate that children attend those programs.

Understanding the Elements of School Readiness

Among early childhood researchers, the construct of school readiness is recognized to be multifaceted, with a growing consensus that it spans the domains of linguistic, physical, social, emotional, and cognitive development, as well as academic abilities (including mathematical and scientific thinking, social studies, and the arts) (Snow & Van Hemel, 2008). Studies have consistently shown that young children's attention, self-regulation, and social behaviors are as important as cognitive abilities when seeking predictors of later academic success (Blair & Razza, 2007; Claessens, Duncan, & Engel, 2009; CCSSO, 2011; Duncan et al., 2007; Harvard Center on the Developing Child, 2004; Ladd, Birch, & Buhs, 1999; Raver & Knitzer, 2002; Zhai, Brooks-Gunn & Waldfogel, 2011).[1]

How Do We Measure School Readiness?

Understanding the developmental characteristics of children as they enter school and the types of early experiences that are linked to school success are of great interest to state stakeholders, including educators at the preschool and elementary levels, program developers, early childhood researchers, and parents. Assessing students at the start of kindergarten is one way to help these stakeholders understand children's developmental strengths and limitations. Thus, a growing demand has emerged for reliable and trustworthy assessments that describe what children know and are able to do when they enter kindergarten. In a comprehensive assessment system, this information also can be used as a catalyst to strategically address the preparedness of all children for the challenges of subsequent grades (Saluja, Scott-Little, & Clifford, 2000).

Research suggests that carefully designed kindergarten readiness assessments can be used to support children's development and measure their readiness for standards-based learning (Ackerman & Barnett, 2005; National Research Council, 2009; Shepard, Kagan, & Wurtz, 1998; Stedron & Berger, 2010). Teachers may seek to use results from readiness measures to provide targeted instruction to those who are most at risk of falling behind (SERVE, 2001a, 2001b). Such measures also can be a valuable component in a comprehensive system designed to monitor progress of young children (CCSSO, 2011). According to SERVE (2000a), school readiness assessments should address the following broad goals:
- better inform teachers and parents of the strengths and needs of children and how to best build on these strengths and support these needs;
- assist all children in reaching high expectations;
- help kindergarten teachers move children toward K–12 educational accomplishment;

- enhance instruction and establish system accountability for child readiness;
- inform early childhood and school systems about adjustments needed to improve services; and ease the transition from preschool to kindergarten for all children.

Since early childhood development is uneven and constantly changing, and young children are better able to demonstrate their developmental competencies in naturalistic settings, traditional approaches to testing may not provide a complete portrait of young children's abilities. The National Association of School Psychologists' position paper on early childhood assessment (NASP, 2005) reminds early childhood professionals that young children's development is rapid and variable, and they often have little test-taking experience and short attention spans. All standardized assessment procedures should, therefore, be used cautiously when making educational decisions for young children (Greenwood, Luze, & Carta, 2002; Meisels & Atkins-Burnett, 2000).

In terms of defining and assessing school readiness, of the 22 states with formal definitions of school readiness in place, only nine require assessment of students entering kindergarten (Bush, 2010). Of the 28 states that do not have formal definitions of school readiness, eight require assessment of students entering kindergarten. A report prepared for the National Council of State Legislatures (Stedron & Berger, 2010) found that 25 states had kindergarten assessment procedures in place; of these, 21 required all students to be tested. Four other states indicated that they were developing statewide kindergarten assessment policies, and 21 had no statewide kindergarten assessment policy.

States and districts have taken a number of different approaches to developing school readiness measures. Seven of the 11 states that report using comprehensive measures of school readiness have developed their own tools, while four states chose commercially developed tests (Stedron & Berger, 2010). Each approach has merits and drawbacks, so stakeholders will need to be prepared to weigh tradeoffs when identifying which to choose. Customized assessments can be developed by states or districts so as to reflect a specific curriculum, but care must be taken to ensure the technical adequacy of these instruments so that interpretations drawn from the results are valid. Furthermore, test development is time-consuming and costly, and may require specialized training that district personnel do not have.

General Approaches to Measuring School Readiness
Determination of whether a child has the requisite skills to enter school will utilize assessment methods that are either "direct" or "observation-based."

Direct assessments require the child to respond to various tasks (some familiar, others novel) that are posed by an examiner who often will not be familiar to the child. These tests are most commonly standardized so that the tasks, administration, responses, and scoring are the same for all children. A key advantage to this approach is that direct assessments are associated with higher reliability across children. Direct assessments are often found to be predictive of school achievement and are strongly associated with socio-economic status (Early Childhood Assessment, 2010). They are most useful when evaluating defined cognitive, perceptual, or physical traits.

A limitation to direct assessment is that it is more obtrusive; the assessment "event" is not a part of the child's usual daily routine, and therefore children may feel uncomfortable or anxious during testing. Administration practices may be unfamiliar to the child or the child may not understand how to respond to questions (USED, 2007). Furthermore, few are developed to measure the social and emotional skills of young children (NAEYC, 2003).

Indirect or observation-based methods are often considered to be "ecologically valid" because they present an account of what a child usually does in their daily activities. These methods include a portfolio of the child's work; observations of the child within a classroom, either by the teacher or by another professional; and behavior rating scales that can be completed by parents, caregivers, or teachers (Snow & Van Hemel, 2008).

Portfolios provide an opportunity to view a collection of child-generated work. These samples of work are collected from a child over time so together they can display what a student has learned to do over the course of a year. But they are time-consuming to score and require well-defined rubrics, scoring guides, and training to ensure standardized application of the scoring rubrics across raters. It can be quite challenging to promote and monitor reliability across time with the same rater and across raters (Sattler & Hoge, 2006; USED, 2007).

Classroom observations can provide a wealth of information about a child, and the information provided can be used to plan instruction. Yet, like portfolios, they are time-consuming, requiring the observer to focus on and record specific behaviors during specified time intervals. In addition, these methods usually require extensive observer training, as observer bias constitutes a major source of error that can threaten the validity of data collected and distort observed effect sizes (Briesch & Volpe, 2007; Hoyt, 2000; Sattler & Hoge, 2006).

Any type of behavioral observation must have a defined goal (exactly what is the purpose of the assessment) and a specific focus (which behavior is to be observed and recorded) (Sattler & Hoge, 2006). All observational methods are to be designed with great care so that the underlying assumptions are clear and the behaviors to be observed should be defined precisely and thoroughly. Operational definitions of the behaviors must be clear and explicit, without inferring causality. Both observations and portfolios may be more comfortable for the child, since the assessments are largely unobtrusive and the teacher/ observer may be a familiar figure.

Behavior rating scales are widely used indirect measurement tools. Behavior rating scales have many of the advantages of observations but take less time to complete (Briesch & Volpe, 2007; Merrell, 2001; Snow & Van Hemel, 2008). Rating scales can provide information retrospectively about the child's overall behavior by relying on recollections of someone who knows the child well; the behaviors targeted are often those that occur infrequently (over weeks or months), as well as behaviors seen daily (Riley-Tillman, Kalberer, & Chafouleas, 2005; Snow & Van Hemel, 2008). In an ideal situation, raters will have an informal set of norms against which to compare the observed behavior to the frequency with which that behavior occurs among the child's peers (Sattler & Hoge, 2006).

As with all indirect methods, one of the greatest challenges to using behavior rating scales is that of reliability within and among raters. What has come to be recognized as one of the most robust findings in clinical research is that parents, teachers and others describing the same child often rate behaviors differently (Achenbach, McConaughy, & Howell, 1987). Rater discrepancies have been found in virtually every method of clinical assessment, in samples of informants encompassing diverse ethnic and cultural backgrounds, and in many clinic samples (De Los Reyes & Kazdin, 2004).

Because the raters know the child well, halo effects (i.e., rating the child globally rather than the behavior specified) and generosity errors (i.e., rating a child more favorably than warranted) can contribute to inaccurate ratings (De Los Reyes & Kazdin, 2004; Briesch & Volpe, 2007). Studies by Mathematica Policy Research (2007, 2008) emphasized the need to train teachers and other caregivers who will complete the rating scales so that their observations are not inflated. It may be necessary to conduct periodic monitoring and/or follow-up training of teachers or caregivers to ensure reliability and validity of data if rating scales are to be used over a protracted time (Mathematica Policy Research 2007; Bagnato, et al., 2002a; Bagnato, et al., 2002b).

Maryland's and Ohio's Early Childhood Comprehensive Assessment System

An Introduction to the EC-CAS Partnership for Project Development
The Early Childhood - Comprehensive Assessment System (*EC-CAS*) is being developed by a partnership which includes the Maryland State Department of Education (MSDE) and the Ohio Department of Education (ODE). The Johns Hopkins University, School of Education, Center for Technology in Education (The JHU CTE) and WestEd's Assessment & Standards Development Services, and its Center for Child & Family Services are the sponsored partners who are building the assessment protocols, supporting technology, and professional development structures.

As one of the first states in the country to implement school readiness measures, Maryland has long been a leader in the field of early childhood education. The MSDE has been at the helm of Maryland's success. In 2000, MSDE developed a customized version of the Work Sampling System (WSS), and assessment system which emphasizes observational and portfolio techniques of assessment. Since its inception in 2001, the Maryland Model of School Readiness (MMSR) Kindergarten Assessment was part of a broader effort to streamline an early learning framework built on standards, instructional alignment of prekindergarten with K-12, professional development, and family communications. MSDE's success and vision led the ODE's leadership to develop a state partnership with the MSDE to create a common *EC-CAS* as parts of the ODE's own successful submission of its RTT-ELCG application. Ohio had implemented the Kindergarten Readiness Assessment – Literacy (KRAL) for many years and decided to shift toward an assessment reflecting all developmental domains. Both states were successful in obtaining funds through the third round of the Race to the Top solicitation, which was exclusively earmarked for early childhood education.

As envisioned, this new family of tools is designed to be one assessment system for all children from 36 months through 72 months. This system includes formative assessments (aligned to developmental learning progressions that define expectations at 6-month benchmarks ages 36 through 72 months), and a kindergarten entry assessment, administered at kindergarten entry (median age 63 months). All measures in this system are:

(a) aligned to a continuum of early learning standards that span seven developmental domains (defined below);

(b) supported by a robust system of professional development for teachers and child care providers, as well as school and district administrators;

(c) accompanied by sufficient resources and support for implementation;

(d) connected to state longitudinal data systems to allow for consistent and meaningful reporting at the student, class, school, district, and state levels; and

(e) monitored and evaluated to ensure fidelity, validity, and reliability (Council of Chief State School Officers [CCSSO], 2011; National Early Childhood Accountability Task Force, 2007; Maryland's Early Learning Challenge Grant Application, 2011).

The graph below describes the relationship of formative assessment and the KEA by domains of learning.

DOMAINS	36 mo.						72 mo.
Social Foundations							
English Language							
Math		Formative			KEA		
Physical-Motor			assessment				
Science							
Social Studies							
The Arts (MD only)							
	Formative assessment:				K-Entry		

Measures in this system will be inclusive, adaptive, and capable of identifying areas of proficiency on specific skills in each of the seven domains relative to the child's chronological age. It is designed to be maximally accessible to young children with a wide range of background experiences and developmental needs, and it will administered via technology. To ensure that the intended depth and breadth of each domain is addressed, measures of assessment standards include on-demand assessments, performance assessments, and observational protocols.

The online assessment data system, applying a newly formed Online Reporting System (i.e., dashboard for reporting) integrates results from the multiple measures and is capable of producing reports for state, district, and school-based administrators; teachers; and parents to provide guidance in the creation of an optimal learning plan for each child. Similarly, results

from all assessment components will be used in conjunction with other sources of information to target instruction to meet a child's particular academic, personal/social, or physical needs, as well as to delineate the child's strengths. Preschool and Kindergarten teachers will be trained in the spring and summer of 2014 to administer the measures with fidelity and to use data formatively to improve instruction and provide research-supported, individualized support to children. Safeguards for inter-rater reliability and fidelity of implementation are built into the training through a "certification process" using a simulator and having school-based resource persons available to oversee the implementation.

All components of the new system will be aligned with developmental learning progressions for each of the following domains of development**:**

- social-emotional development,

- physical/motor development,

- language and literacy,

- mathematical thinking, (Note: OH's "Cognitive" domain includes Math and Science)

- scientific thinking,

- social studies; and

- the arts (optional and used in Maryland.)

In addition, all components of the Comprehensive Assessment System (formative measures and KEA) will be vertically articulated to allow for the measurement of growth over time.

This project design incorporates input from State Advisory Committees, comprised of curriculum specialists, early childhood professionals, and other stakeholders. The design, development and implementation is being reviewed by a national Technical Advisory Council (TAC), facilitated by the Council of Chief State School Officers (CCSSO), comprised of developmental psychologists, early childhood experts, and psychometricians. In addition, practitioners have been involved in the content as well as bias and sensitivity review processes.

Developing the New Early Childhood Comprehensive Assessment System (EC-CAS)

This new assessment system is based on constructs common to both states: the existing Maryland Model of School Readiness (MMSR) Kindergarten Assessment using a customized Work Sampling System (WSS), and the State of Ohio's school readiness screening tool for literacy. This section describes specific steps in developing the components of the new combined system:

(A) new Kindergarten Entry Assessment (KEA) as a measure at kindergarten entry;

(B) formative tools that will be developed and administered by teachers, caregivers, service providers, and other key state stakeholder groups;

(C) online data capture and reporting system;

(D) assessment items electronically delivered by touch screen technology;

(E) online professional development and resources.

(F) student progress monitoring (to be developed later in Version 2.0)

Development of the Kindergarten Entry Assessment (KEA)

The new KEA has the following purposes:

- help determine if entering students have the skills and abilities to engage in and benefit from the kindergarten curriculum by demonstrating skills and behaviors which describe end-of-prekindergarten standards;

- identify individual's children's needs and provide necessary supports to children and teachers;

- inform decision-making about teachers' professional development needs; and

- assist teachers in data-driven instructional decision making to meet students' individual needs through targeted intervention;

- inform prior care (e.g., preschool and child care program) stakeholders;

- provide families with information about their children's learning and development.

The KEA Framework ("Blueprint") development follows the typical research-based test development processes to ensure its validity and reliability as a measure of a child's skills and abilities in each of the developmental domains and to yield domain as well as composite scores for each child.

Its development follows a series of steps: (1) framework development; (2) benchmarking and small-scale piloting of item/task prototypes; (3) item/task development; (4) formal pilot testing; (5) field testing; and (6) operational administration in school year 2014-15.

In keeping with guidelines in the Standards *for Educational and Psychological Testing* (AERA, APA, & NCME, 1999), the project endorsed the use of a framework to guide the development of the KEA. This practice ensured that all of the processes associated with these measures—

including design, administration, scoring, and reporting—are grounded in theory, research, and best practice; conducted systematically; and associated with sufficient transparency to allow stakeholders to judge the appropriateness and technical adequacy of emerging measures (Mislevy & Haertel, 2006). In addition, the framework development process served as a guide, enabling alignment among standards/indicators from both states; types of measures, instruments, and instruction, data use, and professional development. In the summer of 2013, the project developed a KEA Blueprint which includes 28 assessment standards across six domains of learning and four additional assessment standards for the Arts, which will only be used in Maryland.

The framework used to develop the KEA includes five core components: (a) detailed information about the content to be assessed; (b) the measurement theory of action underlying the assessment; (c) a description and justification of the item types that will be used; (d) specifications for each item type; and (e) a blueprint that describes the overall design for the item/task pool or test.

Components of the content section of the framework include: 1) the identification of indicators to be assessed by domain, 2) the assessment limits for the chronological benchmark (63 months for the KEA), and 3) the identification of the levels of proficiency.

Test developers must first specify the content or learning expectations for children in ways that precisely identify which standards or indicators will be assessed and how proficiency, developmental readiness, or satisfactory performance will be defined relative to content in each domain (language and literacy, math, science, social studies, personal/social development, physical development, and the arts);

Once standards or indicators are identified, developers must "unpack" the standards to determine what children should know and be able to do at about 63 months of development. State stakeholders at a number of levels were involved in this process, thus providing a forum for discussion which content should be assessed on each measure and how the critical knowledge and skills embodied in each domain at each age level can most accurately and effectively be measured. During this process, test developers defined vertical connections between standards (both within and across domains) which are aligned with the Common Core Standards in terms of the strands as well as essential skills and knowledge within each domain. The assessment items are being developed to measure the essential skills and knowledge associated with each of the 28 (plus 4) assessment standards through three types of assessment modes: selected response, performance task, and observational rubric. This will allow, after its first census administration in the fall of 2014-15, to longitudinally track results of individual students to assessed grades using the newly implemented assessment developed by

the Partnership of Assessment of Readiness in College and Career (PARCC).

Guiding questions to consider during this process included the following:

- What are the states' common priorities for assessment within each domain at kindergarten entry? What are the collective/common behaviors or learning outcomes? Has each state adequately defined the content (knowledge, skills, and abilities) and constructs (cognitive processes) that are the targets for assessment in their standards?
- Which standards/indicators should be assessed on the KEA? Which are better measured with the screening instruments or formative tools?
- For each of the developmental domains, what are the benchmark indicators of proficiency at approximately 63 months? What must a student know and be able to do to be deemed kindergarten-ready?
- How will levels of achievement or performance for each standard or indicator be differentiated?
- What steps will be taken to ensure that the assessment has addressed the full range (depth and breadth) of standards/indicators in each domain?
- How will the states ensure that the system is inclusive of children with a wide range of knowledge, skills, and abilities?

The project's test developers then identified a measurement theory of action, or cognitive model, that describes how expertise or competence in each domain (e.g., personal/ social) can be measured. Generally, test developers seek to measure a child's level of mastery of a particular characteristic (e.g., self-regulation). The measurement theory of action holds that children have different levels of this characteristic, and that, when measured appropriately, their "scores" on this characteristic will fall along a continuum of least to most. This theory of action is closely linked to the content to be assessed *and* to the specific item types that can be used to measure the characteristics of interest, i.e., that can be used to elicit meaningful information (responses) from children about the precise location of their levels of expertise (score or rating) for that characteristic. According to the National Research Council (2001), demonstrating understanding of a theory of action underlying item and test development is an important piece of evidence to support the validity of results that emerge from these assessments.

Guiding questions to support informed decision-making include the following:

- How will each intended outcome be measured? How will the state confirm that its measures address all aspects of the intended outcome?

- What is the theoretical foundation or theory of action for this measurement plan? How is it linked to expertise within or across domains?

- How will the state test the accuracy and appropriateness of its measurement theory of action for the purpose intended? What evidence should be collected?

The project's test developers also determined which item types are developmentally appropriate for children at approximately 63 months of age and are most effective for measuring the content in the domain. Using an evidence-based approach (e.g., see Mislevy & Haertel, 2006; Snow et al., 2010), each item type was explicitly linked to a particular standard or group of standards (within or across domains) to fit the proposed measurement theory of action. Clear links among the targeted content, the measurement theory of action, and the item types had to be evident and justified. The item types and responses to them may be expressed orally or via a performance task; responses may include pointing to a picture, checking behaviors according to a rubric, completing a task, or performing an action. Item templates (also called item shells or prototypes) can be created for each of the item types deemed appropriate for these new assessments. In addition, the project's test developers had to indicate that the items are developmentally appropriate for optional implementation with technology (i.e., tablet or computer).

Specifications for each item type helped by bringing consistency and quality assurance to the ways in which items and/or tasks are being presented, formatted, and used on each assessment. They describe, for each item type, the important criteria or dimensions (e.g., complexity; length; type of response options) that are needed to effectively measure different standards or indicators. Item-type specifications include guidelines for selection of associated stimuli (e.g., level of complexity of text; type of graphics used) and for the administration and scoring of each item type. Reliance on item-type specifications helps to ensure that all subsequent decisions reached during item development by teachers, expert panels, and/or state leaders in diverse locations are guided by the same set of pre-established guidelines. Specifications also may include allowable strategies for developing item types that are maximally accessible to children with special needs.

Guiding questions for the project to consider when writing specifications included the following:

- What are the presentation and/or formatting characteristics that should be prescribed for each item type?

- What assessment practices (e.g., prompting) will be allowed, especially for children with

disabilities or dual language learners?

- What types of stimuli (e.g., pictures; words) will be used?

- How will test developers ensure that universal and appropriate access are included or considered in the development of each item type?

The project is progressing on the development of the KEA items, applying the aforementioned development criteria, and they will be tested as part of a field test in late 2013 in both states engaging more than 200 teachers and 4,000 children.

Formative Assessment Tools

Both states have expressed interest in the use of formative tools that can be used to gauge the progress of learning in young children. Through use of the formative tools, each child's progress along a continuum of typical development within each developmental domain will be monitored and each child's individual learning trajectory tracked over time. In this way, early educators, working with three and four year olds, and kindergarten teachers will have the information needed to individualize learning opportunities and plan for intervention when needed, and that children are on the path to kindergarten readiness.

The formative assessment model proposed is based upon research-supported learning progressions that define and describe the knowledge, skills, and abilities over time for children ages 36 months through 72 months. The formative assessment model will also define specific learning progressions for each developmental domain. Learning progressions describe the pathway that children typically follow as they learn or the sequence in which knowledge and skills develop (Masters & Forster, 1996). They are empirically validated descriptions of how learning typically unfolds within a curricular domain or area of knowledge and skill (Darling-Hammond & Pecheone, 2010).

Measurement Theory of Action. The project partners generated a measurement theory of action to guide development of formative tools. This theory of action describes how learning progresses in young children (by domain) and how growth will be measured. In addition, the theory of action will describe how the measurement method (e.g., observational) will be operationalized to delineate both qualitative distinctions in development and quantified change over time, the target population to be assessed, and how and when the results will be used.

Continua of learning and development, (i.e., learning progressions) will be organized by key constructs within each domain. Exemplars that describe (within the natural

environment or learning context) the progression for which children with adequate support typically follow will be specified as qualitatively distinct developmental steps. A child's current level of development within each item will be determined when teachers identify the qualitative descriptor that best matches the child's displayed knowledge, skills, and behaviors. Levels of development will be thoroughly described in terms of the behaviors, skills, knowledge, or competencies that may be demonstrated.

Development of Item Exemplars. Item exemplars are being developed in all domains, covering the learning and developmental continuum from 36 months through 72 months, which would be appropriate for teachers to complete at the beginning and end of each year of preschool and the beginning and end of kindergarten. The outcome of this work will be item-level continua that express learning and development that children with adequate support are typically observed to display, showing a progression of learning in each of the developmental domains. In addition, a process to provide universal access to all children will define the levels of support teachers can give to children during the administration of the assessment.

Item Development and Administration. Feedback from pilot and field testing will be used to create a comprehensive bank of items that will be made available to teachers and others administering the assessment. Trainers (experienced specialists with expertise in assessing young children and in implementing evaluation protocols) will help field personnel identify key opportunities to use the exemplars, model exemplary administration practices, using the observation rubrics, and show them how to embed the collection of formative assessment data into curriculum, and how to use the formative data process support ongoing intentional curriculum planning and support of learning.

In addition, the formative assessment process, using items associated to different developmental levels, will guide the efforts of the states' special educators who must report on the implementation of outcome system for early intervention and early childhood special education.

Technical Infrastructure for the Early Childhood - Comprehensive Assessment System (EC-CAS)

The EC-CAS will feature a set of linked systems that will measure, monitor, and improve the school readiness of all children. The specific technology and database architecture is being determined through a process of technical requirements analysis and stakeholder verification. Upon completion of this phase, a more detailed document outlining System Requirements and Technical Architecture was developed which provided the basis for a more accurate technical build. This process will allow for the alignment of system requirements with the assessment

and the professional development components.

Below are the primary functions of the EC-CAS that will be delivered and supported using technology:

- Administration and scoring of the kindergarten entry assessment administered in the fall of the kindergarten year;
- Administration and scoring of the Learning Progression formative assessments administered from age three through the end of kindergarten for either periodic or continuous use;
- Data loading from - and export to - each state's longitudinal data system for a set of critical data elements identified during the requirements analysis process;
- Reporting functionality to support data-informed decision making and readiness monitoring at the local, district, and state levels;
- Delivery and support of professional development – including informational websites, embedded procedural facilitators and prompts, direct learning modules, communities of practice, and a simulation-based certification system to determine scorer readiness.

Rationale for Technology Approach

The use of technology for assessing school readiness of young children, particularly using technology-driven direct performance assessments, is relatively new. Bagnato, Neisworth, and Pretti-Frontczak (2010) reviewed 88 of the leading early childhood assessments commercially available. They rated these among a variety of dimensions. Of the 88, just under half (48.6%) were categorized as having medium to high use of technology.

Technology can provide significant advantages over more traditional methods of assessment, including: immediate concept feedback for the child, data tracking for instructional decision making and longitudinal measurement, increased assessment fidelity, and personalized, targeted instructional and developmental recommendations for parents and educators to meet the unique needs of a given child.

There are many considerations, cautions and limitations related to the use of technology for direct assessment of this age-group. At the same time, observation-only systems, the predominant assessment mode for this population, have faced criticism related to validity and reliability, particularly when employed for identifying a "readiness" level. The primary goal of this project is to create a system that incorporates the best of observational and performance-based systems.

Several tenets culled from the research have guided the approach to using technology. These include:

- Observation is still a primary measurement strategy – While still a primary source of children's learning at this stage of their lives, observation must be bolstered by stronger fidelity measures, ongoing inter-rater reliability measurement, and, if applicable, triangulation of data from multiple observers and from multiple sources. This includes creating ways for observers to meet reliability of measurement among each other, for ratings to occur over time, and for artifact evidence.

- Assessment should not be a single point-in-time – The power of observations and portfolio systems is that they allow for an estimation of ability based on repeated looks at the child's behavior and products. Direct assessment can help strengthen and validate these observations, but should be gathered over time to be most effective. Computerized Adaptive Assessment (CAA) allows for multiple ways of administration, with progressing or decreasing difficulty based on prior performance, ultimately providing a more accurate depiction of the child's ability.

- Technology should enable what is not possible without it - Technology should be used to improve the logistics of the assessment administration, and this includes allowing children to demonstrate applied learning in a variety of ways. Authentic, online scenarios can be developed that allow children to experience cause and effect relationships, or to cluster skills together in ways that are not feasible otherwise. At the same time, some skills are easier and more appropriately assessed without technology, through direct measures.

- Technology supports more standardized and comprehensive data collection - Several commercially available observational assessment systems encourage the use of hands-on performance assessment without the use of technology. With these systems, there can be great difficulty in standardizing the process and monitoring fidelity. In addition, true and accurate data collection on a variety of dimensions is often difficult and unfeasible for teachers given other demands on their time. Introducing virtual performance assessments can raise delivery standardization while multiple data streams are being collected – such as latency to respond to the prompt, number of trials to mastery, and overall criterion score.

- Active Media not Passive Media – research has demonstrated children as young as 30 months can use computer-delivered, interactive media to learn a concept as well as children who experienced the instruction face-to-face. Children who passively watched

a video of the concept, however, struggled to demonstrate the same degree of learning. Assessment using direct technology, therefore, should be active rather than passive.

- Accessible Design - Embedded accommodations, adaptations, and accessibility strategies will be employed to support children with special needs, including English language learners and children with disabilities.

Child Performance Assessment

The EC-CAS will employ three core approaches to Child Performance Assessment (CPA):

1) Virtual Performance Assessment (VPA) that utilizes child-friendly technologies, media, and narrative for direct assessment of children on an ongoing basis, and
2) Hands-On Performance Assessment (HPA) consisting of equivalent assessment items that do not require the use of technology for delivery. The results of HPA assessment items can be captured by the teacher and entered into an observational recording system.
3) Kindergarten Readiness Assessment (KEA) consisting of a mix of VPA and HPA items delivered as a mandatory assessment in the fall of the Kindergarten year and based on selected skills from the 63-month benchmarks on the developmental learning continuum.

Virtual Performance Assessment (VPA) – By the administration of the EC-CAS in school year 2014-15, child-friendly technologies, media, and narrative will be employed to create an engaging learning environment that features direct-probing of selected skills in the areas of math, reading, and social skills. The intent of the VPA will be to engage and challenge children through the presentation of scenarios and problems that need to be overcome before the child progresses. While some choice may exist in the scenarios and problems selected, the VPA will likely utilize a guided system of navigation that guarantees targeted skills are probed sufficiently. The VPA will be age-appropriate, child-oriented, and encouraging. Whether moving up or down along the continuum, the VPA's learning environment should continue to engage and motivate.

It is expected, that the VPA will be designed for ongoing use at school, and will be available for children from 36 to 72 months. It is anticipated that the proportion of items captured through VPA (vs. HPA) will increase with age.

Hands-On Performance Assessment (HPA) – Equivalent assessment items that do not require the use of technology to deliver will be provided for all skills across the learning progressions. The results of these non-technology-based assessments will be recorded by teachers and

entered into the Observational Reporting System (ORS) described below.

<u>The Kindergarten Readiness Assessment (KEA)</u> - In the fall of the kindergarten year, a mandatory delivery of the KEA will occur, based on KEA Framework. Access to these items is being restricted to the KEA. The KEA will have a defined window of availability for completion which is currently set as the first day of school through the end of October. The VPA is expected to include only a subset of the full-expectation of skills at the associated benchmarks. Teachers will also record their observations on the end-user device, i.e., tablets or computer. Schools or early childhood programs without WiFi capability can administer HPA items, i.e., assessment kit, as an alternative.

<u>Observation Reporting System</u>

The Observation Reporting System (ORS) is the interface by which educators will record ratings for a child's performance on the continuum of skills. The ORS will include specific examples of what performance might look like for a given skill before prompting the teacher or parent to enter a rating for the child. Ideally, there will also be a simple way to capture and associate artifact-based evidence for a given observation (audio, video, photos, etc). Each child will be associated with a Teacher of Record (TOR) who determines the official rating, though other caregivers or guardians could potentially enter ratings as well.

Vision for the Professional Development (PD)

A system of professional development is being created that provides sufficient resources and supports for implementation of the screening instruments as well as the formative assessments and Kindergarten Entry Assessment (KEA).

For the statewide training in the spring and summer of 2014, State-approved trainers will rely on a collection of online professional development resources and blended online and face-to-face approaches that promote development of the knowledge and skills needed to use all system components effectively. These approaches are conceived as a collection of learning experiences in multiple formats that provide tiered support for professionals with varying levels of experience in child assessment. The range of professional development activities are designed to develop skills in collecting, interpreting, and using data among school and program leaders, teachers, and parents and to support the development of research-based tools and resources that address emerging needs.

<u>Stages of Assessment</u>

Overall, professional development activities will be organized around three stages of assessment:

- <u>Pre-administration.</u> Professional development related to pre-administration will focus on ensuring that users understand the purpose of the various assessment tools, are thoroughly knowledgeable about issues related to data security and integrity, and know how to effectively communicate with parents and other stakeholders, including administrators, about the purposes and results of the assessments.

- <u>Administration of assessments.</u> Professional development related to administration of the assessments will provide an overview of performance assessments (if applicable), develop understanding of the processes and procedures for each type of assessment instrument, afford opportunities for hands-on use of specific assessment tools and associated resources, promote understanding of accommodations and adaptations for various high-need populations, develop the skills needed to interpret and score children's responses, introduce participants to the data collection and reporting system, and offer opportunities for hands-on use of the system.

- <u>Post-administration analysis and use of data</u>**.** A third set of professional development offerings will focus on the post-assessment analysis and use of data. These materials will focus on increasing understanding of assessment scores, communicating assessment results to parents and families, utilizing data to make instructional decisions and individualize instruction, and providing additional information on data quality and integrity.

The professional development outcomes throughout the stages will:

- Inform audiences of the connection between the formative assessments and the KEA (pre-administration).
- Inform audiences of the purpose and value of the formative assessments and the KEA (pre-administration).
- Inform audiences of issues related to assessing children, data security, integrity, and reliability/validity (pre-administration and post-administration).
- Inform audiences of the importance of effectively communicating the purposes and results of assessments (pre-administration).
- Promote audience understanding of child development (pre-administration):

- o Understand variations in patterns of child development, including the development of young children with disabilities and dual language learners.
- Facilitate audience ability to conduct assessments, including (administration):
  - o Processes and procedures for conducting assessments including capturing relevant artifacts;
  - o Training that incorporates hands-on use of assessment tools;
  - o Understanding accommodations and adaptations for high needs populations;
  - o Ability to score responses with high reliability.
- Facilitate audience ability to view and interpret assessment results (post-administration).
- Support effective data-driven instructional practices to help educators to individualize instruction and move children along the developmental continuum (post-administration).
- Show educators the alignment of the assessments to their existing curriculum (post-administration).
- Communicate specific assessment results to families and stakeholders (post-administration).

Professional Development (PD) Approaches

A variety of professional development approaches will be used and customized throughout the project to accommodate the needs of differing audiences and settings. The range of PD approaches includes:

- Train-the-trainer model
  - o Face-to-face workshops, webinars, and hands-on experience with assessment measures.
  - o Delivered regionally or accessed through online training module with end-of-unit testing.
  - o Follow-up training includes self-paced online modules with built-in capacity to track and certify participants' completion.
  - o Opportunities for interaction with professional colleagues and mentors.

- Web-based training, coaching, and technical assistance
  - o Online tutorials, grab 'n go resources, and mini-lessons that provide individualized support to child care providers and educators.

- Establishment of online learning communities
  - Password-protected, user-friendly online environment for the various audiences.
  - Provides areas for content delivery, collaboration and file sharing.
  - Used to connect and provide sustainable support to audiences.
  - Customizable to include separate communities for different audiences or space to share information and resources across audiences.
  - Includes a repository of state-developed and -vetted resources (e.g., web-based learning modules and tutorials) for improving professional skills and practices and a forum for sharing knowledge, insights, and observations.
  - Examples of resources and online activities include: recommended readings, focused group discussions, and sharing of annotated examples of best practices and exercises to help educators develop expertise within the context of local practice.

- Use of simulation technology
  - Online training approach that provides "real life" hands-on experience and practice for administering assessments and analyzing data for instructional improvement.
  - Used to enhance the inter-rater agreement as the basis for an assessor certification process.

- Integration of an evidence-based Procedural Facilitator
  - Guides users through administering the assessment and provides supports to make informed decisions when entering and interacting with data.

- Development of a web-based portal
  - A custom web portal created for all audiences and stakeholder groups.
  - Allows access to online assessment tools, professional development, and related resources.
  - Integrates cutting-edge social media features for collaboration and resource sharing.

Next Steps

The current work by the partners is being expanded to other states. In September 2013, Maryland, as the lead agent for a KEA State Consortium, was awarded a federally-funded Enhanced Assessment Grant. The participating states on the Consortium are Connecticut, Indiana, Massachusetts, Michigan, and Nevada. The Consortium's intent is to enhance the

current KEA as described here in terms of extending the learning progressions downward to birth, by developing more assessment items, and to strengthen the technology capabilities of the system. The enhanced KEA Version 2.0 is scheduled to be implemented in school year 2016-17.

References

Achenbach, T.M., McConaughy, S.H., & Howell, C.T. (1987). Child/ adolescent behavioral and emotional problems: Implications of cross-informant correlations for situational specificity. *Psychological Bulletin, 101*, 213–232.

Ackerman, D. J., & Barnett, W. S. (2005). *Prepared for kindergarten: What does "readiness" mean?* New Brunswick, NJ: National Institute for Early Education Research.

American Educational Research Association (AERA), American Psychological Association (APA), and National Council on Measurement in Education (NCME). (1999). *Standards for educational and psychological testing*. Washington, DC: Authors.

Bagnato, S.J., Smith-Jones, J., McComb, G., & Cook-Kilroy, J. (2002a). Quality early learning— Key to school success: A first-phase 3-year program evaluation research report for Pittsburgh's Early Childhood Initiative (ECI). Pittsburgh, PA: SPECS Program Evaluation Research Team.

Bagnato, S.J., Suen, H., Brickley, D., Jones, J., & Dettore, E. (2002b). Child developmental impact of Pittsburgh's Early Childhood Initiative (ECI) in high-risk communities: First-phase authentic evaluation research. *Early Childhood Research Quarterly, 17*(4), 559-589.

Blair, C., & Razza, R. A. (2007). Relating effortful control, executive function, and false belief understanding to emerging math and literacy ability in kindergarten. *Child Development,78*, 647-663.

Briesch, A. M. & Volpe, R. J. (2007). Important considerations in the selection of progress-monitoring measures for classroom behaviors. *School Psychology Forum: Research in Practice, 1*(2), 59-74.

Bush, M. (2010, November). *State characteristics: kindergarten; state notes: early learning— kindergarten.* Denver, CO: Education Commission of the States.

Claessens, A., Duncan, G., & Engel, M. (2009). Kindergarten skills and fifth-grade achievement: Evidence from the ECLS-K. *Economics of Education Review, 28,* 415–427.

Council of Chief State School Officers (CCSSO). (2011). *Moving forward with kindergarten readiness efforts.* Washington, DC: Author.

Daily, S., Burkhauser, M., & Halle, T. (2010). A review of school readiness practices in the states:

Early learning guidelines and assessments. *Child Trends: Early Childhood Highlights, 1*(3). Retrieved from http://www.childtrends.org/Files/Child_Trends-2010_06_18_ECH_SchoolReadiness.pdf

Darling-Hammond, L., & Pecheone, R. (2010). *Developing an internationally comparable balanced assessment system that supports high-quality learning.* Presentation at the National Conference on Next Generation K–12 Assessment Systems, Center for K–12 Assessment & Performance Management, with the Education Commission of the States (ECS) and the Council of Great City Schools (CGCS), Washington, DC. Retrieved from http://www.k12center.org/rsc/pdf/Darling-HammondPechoneSystemModel.pdf

De Los Reyes, A. & Kazdin, A. E. (2004). Measuring informant discrepancies in clinical child research. *Psychological Assessment, 16*, 330– 334.

Duncan, G. J., Dowsett, C. J., Claessens, A., Magnuson, K., Huston, A. C., Klebanov, P., Pagani,

Graue, M. E. (2006). The answer is readiness—Now what is the question? Early Education and Development, 17(1), 43–54. Retrieved from http://www.leaonline.com/doi/pdf/10.1207/s15566935eed1701_3

Greenwood, C. R., Luze, G. J., & Carta, J. J. (2002). Best practices in assessment of intervention results with infants and toddlers. In A. Thomas & J. Grimes (Eds.), Best practices in school psychology IV (pp. 1219–1230). Bethesda, MD: National Association of School Psychologists.

Harvard University Center on the Developing Child. (2004). Young children develop in an environment of relationships: Working paper #1. Boston: National Scientific Council on the Developing Child.

Hoyt, William T. (2000) Rater bias in psychological research: When is it a problem and what can we do about it? Psychological Methods, 5(1), 64-86.

Ladd, G. W., Birch, S. H., & Buhs, E. S. (1999). Children's social and scholastic lives in kindergarten: Related spheres of influence? Child Development, 70(6), 1373-1400.

Masters, G., & Forster, M. (1996). Progress maps, assessment resource kit. Melbourne, Australia: Australian Council for Educational Research.

Mathematica Policy Research. (2007). Measuring children's progress from preschool through third grade. Princeton, NJ: Author.

Meisels, S. J., & Atkins-Burnett, S. (2000). The elements of early childhood assessment. In J. P. Shonkoff & S. J. Meisels (Eds.), Handbook of early childhood intervention (2nd ed.) (pp. 231–257). New York: Cambridge University Press.

Merrell, K. W. (2001). Assessment of children's social skills: Recent developments, best practices, and new directions. Exceptionality, 9(1&2), 3–18.

Mislevy, R., & Haertel, G. (2006). *Implications of evidence-centered design for educational testing (Draft PADI Technical Report 17).* Menlo Park, CA: SRI International. Retrieved from http://padi.sri.com/downloads/TR17_EMIP.pdf

National Association for the Education of Young Children (NAEYC). (2003). Early childhood curriculum assessment, and program evaluation. Joint Position Statement of the NAEYC and the National Association of Early Childhood Specialists. Retrieved from http://www.naeyc.org/files/naeyc/file/positions/pscape.pdf

National Association of School Psychologists (NASP). (2005). Position Statement on Early Childhood Assessment. Retrieved from http://www.nasponline.org/about_nasp/pospaper_eca.aspx

National Early Childhood Accountability Task Force. (2007). *Taking stock: Assessing and improving early childhood learning and program quality*. Retrieved from http://www.pewtrusts.org/uploadedFiles/wwwpewtrustsorg/Reports/Pre-k_education/task_force_report1.pdf

National Governors Association (NGA). (2005). Building the foundation for bright futures: Final report of the Task Force on School Readiness. Retrieved from http://www.nga.org/cms/home/nga-center-for-best-practices/center-publications/page-edu-publications/col2-content/main-content-list/building-the¬foundation-for-brig.html

National Research Council. (2009). Mathematics learning in early childhood: Paths toward excellence and equity. Committee on Early Childhood Mathematics, C. T. Cross, T. A. Woods, and H. Schweintruber (Eds.). Center for Education, Division of Behavioral and Social Services and Education. Washington, DC: The National Academy Press.

National Research Council. (2001). *Knowing what students know: The science and design of educational assessment*. Washington, DC: National Academy Press.

Pianta, R. C., Barnett, W. S., Burchinal, M., & Thornburg, K. R. (2009). The effects of preschool education: What we know, how public policy is or is not aligned with the evidence base, and what we need to know. *Psychological Science in the Public Interest, 10*, 49–88. doi:10.1177/1529100610381908

Raver, C.C., & Knitzer, J. (2002). Promoting the emotional well-being of children and families. Policy Paper No. 3. Ready to Enter: What Research Tells Policymakers about Strategies to Promote Social and Emotional School Readiness among Three-and Four-Year-Old

Children. National Center for Children in Poverty.

Rhode Island KIDS COUNT (2005, February). Getting ready: Findings from the National School Readiness Indicators Initiative: a 17-state partnership. Prepared by KIDS COUNT Rhode Island. Retrieved from: http://www.gettingready.org

Riley-Tillman, T. C., Kalberer, S. M., & Chafouleas, S.M. (2005). Selecting the right tool for the job: A review of behavior monitoring tools used to assess student response-to¬intervention. The California School Psychologist, 10, pp. 81-91.

Saluja, G., Scott-Little, C., & Clifford, R. (2000). Readiness for school: A survey of state policies and definitions. Early Childhood Research & Practice, 2(2).

Sattler, J. M. & Hoge, R. D. (2006). Assessment of Children: Behavioral, Social, and Clinical Foundations, 5th Edition. San Diego, CA: J. M. Sattler.

SERVE. (2000). School readiness in North Carolina: Strategies for defining, measuring, and promoting success for all children. Paper submitted to the North Carolina State Board of Education. Greensboro, NC: SERVE, University of North Carolina, Greensboro.

SERVE. (2001a). Assessing kindergarten children: What school systems need to know. Greensboro, NC: SERVE, University of North Carolina, Greensboro.

SERVE. (2001b). Making them work for all children. Paper Presented at SERVE Institute on School Readiness Assessment Systems: Making Them Work for All Children. Retrieved from http://oldweb.naeyc.org/ece/critical/readiness.asp

Shepard, L., Kagan, S.L, & Wurtz, E. (1998). Principles and recommendations for early childhood assessments. Washington, DC: National Education Goals Panel. Retrieved from www.negp.gov/reports/prinrec.pdf.

Snow, E., Fulkerson, D., Feng, M., Nichols, P., Mislevy, R., & Haertel, G. (2010). *Leveraging evidence-centered design in large-scale test development*. Large-Scale Assessment Technical Report 4. Menlo Park, CA: SRI International.

Snow, K. L., & Van Hemel, S. B. (2008). *Early childhood assessment: Why, what and how?* Washington, DC: National Research Council.

Snow, K. L. (2006). Measuring school readiness: Conceptual and practical considerations. Early Education and Development, 17(1), 7–41.

Stedron, J., & Berger, A. (2010). *NCSL Technical Report: State approaches to school readiness assessment.* Denver, CO: National Conference of State Legislators.

U.S. Department of Education (USED). (2007). A review of methods and instruments used in state and local school readiness evaluations (Issues & Answers Report, REL 2007–No. 004). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory–Southeast. Retrieved from Retrieved from http://ies.ed.gov/ncee/edlabs

Zhai, F., Brooks-Gunn, J., & Waldfogel, J. (2011). Head Start and urban children's school readiness: A birth cohort study in 18 cities. Developmental Psychology, 47, 134–152.

Zins, J.E. & Elias, M.J. (2006). Social and emotional learning. In G. G.Bear & K. M.Minke (Eds.), Children's needs III: Development, prevention, and intervention (pp. 1–13). Bethesda, MD: National Association of School Psychologists.