Developing Kindergarten Readiness and Other Large-Scale Assessment Systems

NECESSARY CONSIDERATIONS IN THE ASSESSMENT OF YOUNG CHILDREN

by Kyle Snow, PhD

December 2011

naeyc®

RESEARCH POLICY PRACTICE

naeyc®

National Association for the Education of Young Children
1313 L Street NW, Suite 500
Washington, DC 20005-4101
202-232-8777 800-424-2460
www.naeyc.org

Developing Kindergarten Readiness and Other Large-Scale Assessment Systems

Necessary Considerations in the Assessment of Young Children

Over the past few years, interest in assessing children as they enter kindergarten has gained momentum in states. Roughly half of the states have instituted some form of kindergarten entry or readiness assessment, even before the development of such an assessment was highlighted in the Race to the Top, Early Learning Challenge funding competition. The development of readiness assessments has varied across states in terms of the areas of child development and knowledge evaluated, and their use for policy and practice purposes. The Center for Applied Research at the National Association for the Education of Young Children (NAEYC) has developed this guidance to support states' development and implementation of kindergarten readiness assessment systems. Such systems, properly developed and implemented, can contribute greatly to the success of early childhood programs and early elementary programming to identify and meet the needs of children entering kindergarten. The considerations presented in this paper are built around NAEYC positions related to assessment and research on child assessment. While focused on large-scale assessment of young children, the guidelines included here are intended to inform considerations about early childhood assessment beyond the implementation of kindergarten entry assessments.

Suggested citation:

Snow, K. 2011. Developing Kindergarten Readiness and Other Large-Scale Assessment Systems: Necessary Considerations in the Assessment of Young Children. Washington, DC: National Association for the Education of Young Children.

Kyle Snow, PhD, is Senior Scholar and Director of the Center for Applied Research at the National Association for the Education of Young Children.

Launched in 2010, the Center for Applied Research encourages and supports communication about research in early childhood development and education among practitioners, policy makers, and researchers. The Center facilitates dialogue through interpretation and translation of research to inform practice and policy, and by identifying and seeking to meet the research needs of practitioners and policy makers in early childhood.

Developing Kindergarten Readiness and Other Large-Scale Assessment Systems

Necessary Considerations in the Assessment of Young Children

There has been much more attention in recent years to the importance of high-quality early childhood programs that provide children with experiences that nurture their development and prepare them for success in school and beyond. While much attention is paid to ensuring the quality of early childhood programs, there is an increasing focus on the role of assessment within early childhood systems, not just on the use of assessment by programs for improving teacher strategies and services in the classroom. While there is broad consensus that early childhood assessment can play a vital role in improving instruction within the classroom, how assessment of young children can and should—and should not—be used to determine program effectiveness is more contentious. NAEYC believes that there is an appropriate role for information from child assessment in large-scale system efforts when attention is also given to research on child development; other indicators that impact children's development and learning; and, best practices in the field as well as assessment science to guide the development, implementation, and use of assessment systems.

NAEYC believes that there is an appropriate role for information from child assessment in large-scale system efforts when attention is also given to research on child development; other indicators that impact children's development and learning; and, best practices in the field as well as assessment science to guide the development, implementation, and use of assessment systems.

In the last decade, policy discussions of assessment in early childhood systems have grown, reflecting the increasing demand for accountability in the elementary and secondary public education system as well as increased state funds for prekindergarten programs. Assessment in early childhood is not a recent concern (see Meisels 2007). Over the past few decades, assessment of young children has been attempted, often with unintended negative

consequences (see Shepard 1994, for a review). Earlier efforts saw kindergarten readiness screening as a means of identifying children deemed ready for school and tracking those not ready into alternative programs, or denying access altogether. However, recent advances in theory have connected assessment with child learning, making assessment part of the "learning culture" (Shepard 2000). While there have still been missteps—efforts to launch the Head Start National Reporting System, perhaps (Meisels & Atkins-Burnett 2004)—the use of assessments for young children continues to be seen as an important element of early childhood programs. Over the past few years, interest in assessing children as they enter kindergarten has gained momentum in states. Roughly half of the states have instituted some form of kindergarten entry or readiness assessment (Daily, Burkhauser, & Halle 2010; Stedron & Berger 2010). They vary in scale of children assessed, the areas of child development and knowledge that are evaluated, and in use for policy and practice purposes.

State interest in the use of kindergarten readiness assessments is growing in terms of the number of states as well as evolving in terms of the implementation and use of such assessment information.

State interest in the use of kindergarten readiness assessments is growing in terms of the number of states as well as evolving in terms of the implementation and use of such assessment information. This year, development of comprehensive assessment systems, including kindergarten entry assessments, is a criterion in the federal Race to the Top, Early Learning Challenge application.

Other efforts at determining children's "readiness" have gone beyond an assessment at entry to kindergarten or pre- and post-assessment in the prekindergarten year. Seventeen states participated in the School Readiness Indicators Initiative that set state indicators for children's progress from birth through age 8 to guide state policies. These indicators look not only to children's progress, but also the context of ready families, ready communities, and ready schools and their impact on children's development. Notably Pennsylvania and the District of Columbia are gathering information on the context in which children develop to gain a fuller picture of children's readiness for kindergarten beyond a child assessment data field. Work in Canada in developing the Early Development Instrument (EDI) (Janus et al. 2007) provides an example of an intentionally designed, population-based assessment of young children's learning across multiple domains. Subsequent development around the EDI addresses how data may (and may not) be effectively used within communities to gain a broad picture of children within the community (see, e.g., Guhn, Janus, & Hertzman 2007).

NAEYC'S ROLE IN GUIDANCE ON CHILD ASSESSMENT

The National Association for the Education of Young Children (NAEYC) has developed this guidance to support states' development and implementation of kindergarten readiness assessment systems. However, principles in developing a statewide assessment program for young children are applicable for other

large-system users, as well as individual programs interested in implementing a common assessment strategy across classrooms and/or programs. This paper underscores that assessment as children enter kindergarten exists within the larger framework of efforts to improve both the birth-to-kindergarten span and the early elementary grades experiences of children in their respective states. Viewing assessment in the context of the birth through third grade continuum and in the context of a comprehensive assessment system underscores how decisions made at one juncture influence the assessment system and the education system elsewhere. In addition to offering considerations from a policy perspective, this paper also extends to implementation practices.

This paper underscores that assessment as children enter kindergarten exists within the larger framework of efforts to improve both the birth-to-kindergarten span and the early elementary grades experiences of children in their respective states.

NAEYC has long promoted the use of developmentally (including culturally and linguistically) appropriate assessments of young children to improve instruction and programs. NAEYC's early childhood program accreditation system is framed by 10 standards of program quality, one of which is assessment of child progress.[1] The NAEYC Standards for Teacher Preparation (teaching children from birth through age 8) for associate, baccalaureate, and higher degrees also address teachers' understanding and use of child assessments.[2]

The National Association for the Education of Young Children (NAEYC) and National Association of Early Childhood Specialists in State Departments of Education (NAECS/SDE) (2003) position statement on curriculum, instruction, and assessment underscores the need for a systemic, integrated approach for linking assessments with curriculum and instruction. This approach was reiterated and furthered by the National Academy of Science panel on early childhood assessment (Snow & Van Hemel 2008, p.10), which stated explicitly "... that a primary purpose of assessing children or classrooms is to improve the quality of early childhood care and education by identifying where more support, professional development, or funding is needed and by providing classroom personnel tools to track children's growth and adjust instruction." As such, aligning assessments with curriculum and instruction ensures that the intended outcomes are addressed and monitored, while misaligned systems cannot ensure that intentions (through standards, curriculum, and instruction) are being met (Martone & Sireci 2009). The greatest danger is an assumption that alignment exists, and making decisions based upon that assumption, when in fact there is little or no alignment. Achieving alignment presents special challenges in early education. While most states have early learning guidelines and standards for K–12 education systems, some do not, and those that do vary (Scott-Little et al. 2007). Within states, there is variation in the curricula used by programs. These variations mean that aligning an assessment system with state standards may or may not also mean that the assessment is aligned with locally implemented curriculum.

THE STATE OF EARLY CHILDHOOD, INCLUDING KINDERGARTEN

Roughly 12 million children in the United States experience some out-of-home settings prior to kindergarten. They may be in formal child care centers, family child care, relative/neighbor care, Head Start or Early Head Start programs, or school settings. The early care and education system reflects the diversity of parents' needs and desires for nonparental care for their children during the years before school entry. Children may attend these programs as young as 6 weeks old and as much as a full work week in order to meet their parents' work needs. Many children are in programs in the evening while their parents are employed in shift work. The quality of settings across the early care and education spectrum varies widely, as do the reasons parents' choose specific programs and their expectations for the types of experiences programs provide to children.

With the exception of Head Start and Early Head Start, requirements for program standards and early learning content standards are designed and implemented by states. Wide variation exists among states in terms of basic child care licensing standards from ratios and group size to teacher qualifications and professional development. States also differ in their standards for state-funded prekindergarten programs. Even though every state has early learning standards for preschool children, and a majority now have such standards for infants and toddlers, states do not mandate their use in all settings. Many children also spend numerous hours a week with a legally unregulated caregiver or program exempt from child care licensing standards. This variability in policies and standards across states, and sometimes even within states, creates an early childhood education sector with a broad range of quality and abilities to meet young children's developmental needs.

High-quality programs are also frequently seen as a means of meeting the needs of children born into poverty. Still, only one in six eligible families receives public assistance to afford child care, less than half of the eligible preschool children can enroll in Head Start, and fewer than 5 percent of eligible infants and toddlers enroll in Early Head Start due to significant lack of public resources to meet the needs of children.[3] Despite the apparent scope of federal funding, it is not sufficient to provide access to high-quality care for all young children. While there has been momentum in the states to fund public prekindergarten, some states have reduced their investment and/or lowered quality standards in the last year, and program quality varies both across and within states. This poses additional questions on the design, implementation, and use of child assessments given the varied requirements that lead to quality for children in a single day (half-day Head Start; half-day, state-funded prekindergarten with a child care subsidy for the full day and full year) let alone a span of five years before enrollment in kindergarten.

Kindergarten readiness also does not have a common, national definition (Snow 2006). Kindergarten policies and attendance also vary widely across the country. While 45 states or territories require school districts to offer kindergarten, only 19 states or territories require children to attend kindergarten prior to

entering first grade. Cut-off birthdates for kindergarten entry eligibility vary among the states, creating differences in age at entry, meaning that children enrolled in kindergarten for the first time across the country can range from younger than 5 years old to nearly 7 years old.[4] Only 10 states require school districts to provide full-day kindergarten; 34 states and the District of Columbia require provision of half-day kindergarten. The variations in requirements for kindergarten teachers' preparation and licensure across states mean that teachers of kindergarten children may or may not have specialized early childhood education degrees, certificates, or endorsements. Across states, the kindergarten year exists within a space somewhere between early childhood, for which most states have early childhood learning standards, and K–12 systems, for which states have standards that may or may not align with those for younger children. Just a handful of states have kindergarten standards that address children's social and emotional development and approaches to learning, in addition to cognitive content areas and physical development.

THE STATE OF ASSESSMENT OF YOUNG CHILDREN[5]

The current state of assessment in early childhood remains, as stated by the National Academy of Science's *Eager to Learn: Educating Our Preschoolers,* "in flux" (Bowman, Donovan, & Burns 2001). In its follow up report in 2008, the NAS noted that there is a great need for additional research and development of assessments appropriate for young children (Snow & Van Hemel 2008). Specifically, the report indicated the need for more and better assessments across all developmental domains, especially domains of social and emotional functioning and development. In addition, more and better assessments are needed that are appropriate for use with infants and toddlers, for dual language learners, and for children in the early elementary grades. Throughout, the NAS urged caution about the inappropriate uses of assessments of young children, a caution that remains highly relevant for policy makers and practitioners.

A one-time snapshot of a child entering a kindergarten classroom cannot capture all of the cumulative experiences in programs, in the home, and in the community of a young child from birth to that day in kindergarten.

A one-time snapshot of a child entering a kindergarten classroom cannot capture all of the cumulative experiences in programs, in the home, and in the community of a young child from birth to that day in kindergarten. Such assessments should not be seen as reflecting on the quality of early care and education during the prekindergarten year in isolation from demographic risk, experiences in the home and the community, other early care and education experiences, and the resources available to support professional development and improve quality. That said, assessments should be done throughout the kindergarten year to help the teacher target and recalibrate his or her efforts over time. The entry assessment provides the teacher with baseline information.

The statute authorizing the Race to the Top, Early Learning Challenge (ELC) grants requires states that develop kindergarten readiness assessments to comply

in their use of child assessments with the reports of the National Academy of Science. Those reports outline the framework of a system of assessment, the resources needed at the system and classroom levels to select and use assessment of young children to improve instruction and services, and the cautions of inappropriate use of assessments of children, particularly when the early childhood system as a whole has so much fragility and variability especially for vulnerable children and families.

ASSESSMENT WITHIN EARLY CHILDHOOD PROGRAMS

Integrating assessment into early care and education programs (those serving children from prekindergarten through third-grade) also requires that resources be allocated to the assessment itself, above and beyond what is necessary to fund the assessment programs (costs of staff, materials, etc.) as well as the delivery of services provided by the early childhood program. With finite financial resources, states and program developers must balance the expense of assessment against other demands for the financial resources. To that end, this paper is intended to provide guidance about implementation of assessment systems and use of data they provide, as well as to provide a rationale for the importance of a well-developed system in most effectively developing and implementing early childhood education systems and programs.

BASICS OF ASSESSMENT

Many of the participants in planning assessment systems as called for by the ELC are cognizant of basic assessment principles. However, some may not be, and even for those with a background, knowledge, or expertise in assessment, it is useful to pause to ensure that all are using a common vocabulary and set of assumptions.

The quality of an assessment is a direct function of its psychometric characteristics

The best measure of the quality of an assessment is built around the degree to which it meets or exceeds standards for various psychometric qualities (see AERA, APA, & NCME 1999). The most basic of these properties are reliability and validity. Reliability refers to the degree to which an assessment provides the same result when administered by different people to the same child, or to the same child at two time points in close proximity. Validity refers to the degree to which the results of the assessment accurately capture what they are intending to capture.[6] Publishers or developers of existing assessments should be able to provide psychometric information on the assessment. This information should also include a description of the nature of children (i.e., the sample) upon which the information is based.

The best measure of the quality of an assessment is built around the degree to which it meets or exceeds standards for various psychometric qualities.

For norm-based assessments, this may or may not also be the "norming sample," the sample around which the assessment's norms were developed. Assessments being used with children dramatically different from those used by the developer to establish the assessment's psychometric properties may not produce the same degree of soundness. If an assessment's psychometrics are not available for children comparable to those being assessed, or if the assessment being used is one being newly developed, psychometric data need to be collected and evaluated before the assessment is adopted and used widely.

There are different types of assessments

For older children, the preponderance of assessment is done in a large group setting, with paper and pencil (and increasingly computer based) formats. This approach is not suited for young children. Assessment of young children, as a result, can be of many types and formats that differ primarily on the source of information. Direct assessments are those administered directly to the child. Direct assessment can be costly, time consuming, and require specialized skills to administer; however, because responses are coming directly from the child they are also often assumed to be objective and accurate (although they may in fact vary in validity). Observational measures do not require explicit administration to children, as they are typically assessments completed by adults after or during a period of observing the child. Observational measures may include rating scales or checklists of specific skills that are demonstrated during the observational period. Finally, some observational assessments are conducted in context, such as authentic assessment, where the items are completed within the flow of typical activity. As such, they are often the least obvious or intrusive assessments, and sometimes the least time consuming. They are also, however, more subject to "opportunity to observe" effects—an assessment may call on a rating or report of a specific behavior or skill that does not occur within a given context. While observational methods can provide information about child ability, they may be influenced by other factors related to the observer—how rigorously observations are made and recorded, any assumptions the observer may have about the performance of a specific child or group of children, and various interpretations of the target behaviors or skills.

All assessments are "standardized"

In early childhood, there is great sensitivity to the idea of "standardized assessment." The popular concern is that of very young children completing paper and pencil assessments en masse, similar to perceptions of large-scale standardized assessments used for older children. The use of this type of assessment is not appropriate for young children. However, the concept of standardization is relevant. Briefly, standardization means that an assessment is administered in the same way, each time it is administered (regardless of who is delivering and who is responding to the assessment). Violating standardization certainly undermines the assessment's reliability, but also threatens its validity. For example, an item on a direct mathematics assessment may have been developed to allow children to use some manipulable (e.g., counters, blocks) to help solve a basic addition problem. When conducting this assessment, all

children should have access to the appropriate manipulable. Likewise, in a direct assessment, even if in the estimation of the assessor the child guesses the correct response, the response given is the one accepted, and the scoring protocol for the assessment would likely (if it is psychometrically sound) account for some proportion of correct guesses. Assessments that are conducted "in context," such as authentic assessments, may come close to violating standardization if they are completed in different contexts for some children than others. For example, authentic assessment of reading done for most children during "reading group" but for some children during a session of targeted individual literacy activity may provide skewed results for these children.

With these general concepts as background, it is important to place the assessment itself into a larger system context. As described below, the assessment context will consider some of these characteristics.

Considerations for Large-Scale Assessments

FRAMING AN EARLY CHILDHOOD ASSESSMENT SYSTEM

High-quality early childhood education is supported by assessments aligned with instructional goals and approaches. Assessment, however, does not refer simply to the tool being used; it refers to an interconnected system of decisions and activity. In addition to selecting an assessment tool or tools, the system requires supports and procedures to effectively and appropriately administer the assessment, as well as a data management and analysis system that captures the results of the assessments and allows the data to be used appropriately. These three components—selection, administration, and utilization—collectively comprise the assessment system as it is defined here.

Assessment, however, does not refer simply to the tool being utilized; it refers to an interconnected system of decisions and activity.

The three components of the assessment system are connected and should be strategically and appropriately linked to ensure that the assessment system is supporting the goals not only of the assessment program but of the education program as a whole. Choices made relative to one component will have implications in each of the others. Understanding these implications enables planners to purposefully and intentionally design assessment systems that align with their early childhood programs, which themselves should align with education systems serving older children in the state. Failure to do so creates the risk for misalignment, with potentially unintended consequences (Meisels 1987; Neisworth & Bagnato 2004). For example, a specific assessment may be selected because of its ease of administration, but result in data too limited to inform instruction or evaluate a program's effectiveness. Likewise, an assessment may be administered to a sample of children served, rather than all children, to reduce costs or allow for more in-depth assessment, but these data cannot then be used to inform instruction for all children or for any individual child.

To assist the decision making and planning necessary to deploy the type of assessment system called for by the Early Learning Challenge, this guidance is

organized around the three components of an assessment system—selection, administration, and utilization. For each component, best practices are highlighted and appropriate cautions or caveats provided. It is not the intention of this guidance to advocate for one assessment over another, or even to advocate for one approach over another. This guidance is intended, however, to advocate for taking a systematic approach to assessment. With such an approach, the implications of decisions made within each component for others are identified so that they may be made intentionally by system developers.

ASSESSMENT SELECTION

The current state of the field in early childhood assessment is captured by the National Academy report (Snow & Van Hemel 2008). While there are hundreds of assessment products for young children on the market today, there is unevenness in the degree to which they cover important domains of school readiness, as well as the degree to which they are appropriate for diverse populations, primarily those children who are English language learners (ELLs) or children with disabilities.[7] Available assessments also vary in their quality as measures (reliability and validity). The variation in assessments along these multiple characteristics needs to be carefully considered when selecting specific assessments. However, it should be noted that one option is to develop a new assessment tool, rather than use existing assessment instruments. Addressing this possibility is not specifically included here, but many of the considerations given in selecting an assessment also apply to constructed assessments, as do the implications of these considerations for other aspects of the assessment system.

WHAT IS THE PURPOSE OF THE ASSESSMENT?

The National Education Goals Panel identified four purposes for assessment, including to support instruction, to identify those with special needs, for program evaluation, and for high stakes accountability (Shepard, Kagan, & Wurtz 1998). Assessments are designed to be psychometrically strongest when used for specific purposes. Screening assessments are designed to have their greatest reliability and validity when used to identify children who may be at risk for or experiencing developmental delays and need additional diagnostic follow-up. Using screening instruments to assess child outcomes in other ways (that is, degree of skill) fails to capitalize on the strength of the assessment and may yield non-reliable or even invalid results when used differently. Articulating the purposes of an assessment is necessary for selecting an appropriate assessment tool or tools.

Assessments are designed to be psychometrically strongest when used for specific purposes.

CONSIDERATIONS AND CAVEATS:

- Clearly identifying the purpose of assessment is closely tied to the utilization aspects of the assessment system. Importantly, the design of the assessment is driven by its purpose. How the results of an assessment are used may or may

not be consistent with its purpose and therefore may be inconsistent with its underlying design.

- Using assessments designed for one purpose for another purpose threatens the assessment's psychometric soundness. However, the way in which assessment data are used could be modified to capitalize on the assessment's psychometric strength while still providing important utility. For example, while not an optimal measure for change over time, screening measures can be used to indicate how the proportion of children at risk changes over time (i.e., the percentage of children identified for further diagnostic assessment may be reported over time to describe changes in the size of that group, even though the data cannot be used to measure specific degrees of growth in the same area).
- As states grow comprehensive data systems, data from a range of assessments and other sources of information may be captured increasing the temptation to conduct a broader range of analyses than the data warrant, based upon how they were obtained. Even when analyses are conducted a long time after initial collection, the purpose and best use of assessment data need to be considered before conducting analyses and drawing conclusions from them.
- Use of kindergarten readiness assessments as a means of screening children into or out of kindergarten is inconsistent with generally accepted best practices and NAEYC's formal position on the inappropriate use of kindergarten assessments to keep an age-eligible child from enrolling (NAEYC and National Association of Early Childhood Specialists in State Departments of Education [NAECS/SDE] 2001).
- The purpose of assessment should be clearly communicated to professionals within the education community, policy makers, and children and their families. In addition, the intended uses of the results of assessment should be articulated so that the purpose and utilization components of the assessment system are clearly linked.

Use of kindergarten readiness assessments as a means of screening children into or out of kindergarten is inconsistent with generally accepted best practices and NAEYC's formal position on the inappropriate use of kindergarten assessments to keep an age-eligible child from enrolling.

HOW PSYCHOMETRICALLY SOUND IS THE ASSESSMENT?

As noted above, the best measure of the quality of an assessment is its psychometric properties, chiefly reliability and validity. There are general guidelines and standards available for most psychometric statistics (e.g., AERA, APA, & NCME 1999). Generally speaking, assessments can be judged as either meeting or exceeding these standards or not—modest variations in psychometric statistics across assessments that are otherwise psychometrically sound do not mean a great deal in practice. However, while reliability statistics may be presented along comparable lines (i.e., test-retest reliability assessed two-weeks

apart), variations in how these psychometrics were established create complexity. This is especially likely for validity data, because assessment validity may be reported in a number of ways, oftentimes using different benchmarks for validation. Some validity data may be based upon predictive validity (especially validity data for screening tools), while others may present validation data against any number of other assessments of the same domain of interest. The assessments against which any specific assessment is validated are themselves often of varying psychometric quality—a given assessment may be highly correlated (i.e., valid) with another assessment which may be more or less psychometrically sound itself.

CONSIDERATIONS AND CAVEATS:

- Assessments are primarily judged by their psychometric properties—if psychometric data do not meet or exceed standards then the assessment should not be considered. There are general guidelines and standards available for most psychometric statistics (e.g., AERA, APA, and NCME 1999).
- Psychometric data that are available are often collected under optimal conditions which may or not be achievable during implementation. As a result, they may be somewhat weaker in application. If psychometric data for the population to whom the assessment will be administered are not available, either because the developer does not provide them for the groups of interest or because the assessment is under development, use of the assessment may be considered, but plans for establishing its psychometric properties should be built into the assessment program.
- Psychometric data for two assessments may not be provided in easily comparable ways (i.e., comparing apples with oranges). In such cases, additional psychometric study, consultation with psychometric and content experts, or both, may be necessary to identify the strongest measure.

WHAT IS THE CONTENT OF THE ASSESSMENT?

By definition, assessments provide means of evaluating individuals' skills, abilities, or traits. Identifying what information is of interest is therefore a critical step in selecting an assessment. While the language in the Early Learning Challenge indicates the need for kindergarten readiness assessments, there are continuing debates about the definition of readiness, including what domains should be included and how they should be assessed (see, e.g., Snow 2006). In addition, some domains of interest in early education have a wide selection of assessments currently on the market, while other domains are captured by fewer assessments (e.g., Snow & Van Hemel 2008). Finally, some assessments are designed to provide information across multiple domains. In this case, the specific content of the assessment that contributes to the measurement of domains should be considered in addition to the utility of the assessment as a whole.

CONSIDERATIONS AND CAVEATS:

- Determine what is important to measure first, then find ways of mea uring it, rather than considering what measures currently exist and assume that those are the most important domains to measure. Such a strategy may reinforce the perceived importance of domains because they can be assessed (Kowalski, Brown, & Pretti-Frontczak 2005), even though contemporary research is demonstrating the importance of multiple domains of early childhood development (e.g., Diamond 2010; Snow 2007).

Determine what is important to measure first, then find ways of measuring it, rather than considering what measures currently exist and assume that those are the most important domains to measure.

- Broad spectrum assessments (typically referred to as "readiness scales") may have greater psychometric strength around some domains than others. While these assessments may provide psychometrically adequate or sound "overall" indicators, domain or subscale scores may not be as vigorous.
- Assessment titles may suggest certain content or domain coverage, but the content of the assessment should be examined to ensure that it is capturing what is presumed. For example, the Peabody Picture Vocabulary Test (PPVT) is a measure of receptive vocabulary (i.e., words that a child recognizes only) and not productive vocabulary (i.e., words a child can produce).
- Some domains of interest in early education have a wide selection of assessments currently on the market, while other domains are captured by fewer assessments (see Berry, Bridges, & Zaslow 2004; Bagnato, Neisworth, & Pretti-Frontczak 2010; Boller et al. 2010; Brown et al. 2007; Halle et al. 2011; Snow & Van Hemel 2008 for reviews).
- The content of the assessment should align with state standards and curriculum goals for the children assessed. As noted above, in early childhood, especially in prekindergarten settings, variations in state early learning standards and curriculum use means that an assessment may align with the state standards and be informative for evaluation purposes (e.g., Roach et al. 2010), but may not align closely with locally used curriculum and so have less utility in informing instruction.

WHAT IS THE COST OF THE ASSESSMENT?

The total costs associated with an assessment system can vary dramatically based upon a number of decisions, but are generally high. Depending upon the funding source(s) for adopting an assessment system, the potential drawing of funds away from direct services is a possibility that must be weighed against the ultimate value of the information obtained from the assessment. The costs are spread throughout the system (and will be noted as appropriate). However, with regard to the selection of an assessment, there are a few drivers.

CONSIDERATIONS AND CAVEATS:

- Assessment prices vary considerably in terms of direct purchasing or development costs, as well as the ongoing cost of training (in some cases) and scoring (in some cases). The cost and training demands are also related to issues of administration, as noted below.
- Assessments, like other education resources, typically go through cycles of revision or elimination, which can create recurrent costs as assessments revise over time or need to be replaced.
- Assessments for young children may exist within products lines, typically within publishers or developers. In some cases, there may be vertical alignment (i.e., a series of assessments appropriate to children of different ages that have established psychometric linkages) that may affect costs and/or utility when considered within the larger education system needs.
- Some schools or programs use a curriculum with an integrated or linked assessment. These schools may have an established use history and experience with specific assessments, and the school system may also have paid for use of the materials over a number of years. As a result, the challenges and costs of adding or changing assessments may be an important consideration. Likewise, programs may be under a mandate through their funding source to include specific assessments. In these cases, any additional assessments should be used based upon their capacity to add valuable information about the children being served, either through greater psychometric strength, greater domain coverage, or greater potential to meet additional purposes of assessment.
- Some costs are not so obviously linked to funding money. Time spent preparing for and conducting an assessment that does not ultimately provide information (e.g., the selected tool was not aligned with curriculum or standards) is wasted and needs to be recovered for teachers, program administrators, and children.

ADMINISTRATION

An assessment system must also include consideration of several issues related to the administration of assessment tools—who assesses the child, when, how often, and what degree of training and support is necessary to ensure that administration yields the most psychometrically sound results. Administration also considers which children are assessed, if all children are assessed or a sample of children, if there are assessments in multiple languages and screening for English proficiency.

Who assesses the child?

Assessments for young children include direct assessment instruments, almost exclusively individually administered, or observational measures completed by parents, teachers, or other adults. While these designs have important differences, underlying each is the expectation that there is some relationship between the

child and assessor. Even if the assessor is an adult not previously known by the child, training in administration of early childhood assessment includes rapport-building. This relationship context is important in the direct administration of assessments to encourage the child's participation, but it is also important in observational and rating scale assessments, which often rely upon some experience of the child by the assessor.

The nature of the assessment sometimes determines the assessor; in authentic assessment, the child is assessed by the teacher or caregiver. For many direct assessments, assessors trained specifically on the administration of the assessment are typically used, although it is possible to train teachers on the administration of many measures. However, direct measures (and other measures, as well) require varying amounts of training and experience to administer, making the match between assessor and instrument critical. Determining who will fill the assessor role has consequences for costs, time, and training, as well as implications for data quality.

Determining who will fill the assessor role has consequences for costs, time, and training, as well as implications for data quality.

The purpose of the assessment may also inform the characteristics of the assessor. When the purpose is to inform instruction, the assessment is most typically conducted by the teacher. When the purpose is to measure program effectiveness, the use of assessors other than the child's teacher is often viewed as providing a more objective assessment, although as noted above, it is still important that a relationship exist with the child, even if only through rapport-building. The increasingly cultural and linguistic diversity of children in early education programs raises important questions about the language of assessment (see below). Children who are dual language learners may be assessed in multiple languages, which requires that assessors who are fluent in other languages be able to conduct assessments.

Finally, the assessor should be trained, and able to demonstrate competence, in conducting assessments with young children. In the context of observational measures, this includes consistency of item interpretation and rating criteria, regardless of setting. For direct assessment methods, the need for training includes general procedures in conducting direct child assessments such as rapport-building, avoiding unintended coaching, correct and incorrect use of prompts. For all assessment strategies, training that addresses best practices in "stage setting" (i.e., identifying and using adequate space, minimizing noise and distraction), may also be necessary (e.g., Begeny & Buchanan 2010). Likewise, the importance of conducting the assessment in the same way for all children (see discussion of standardization, above) within a classroom, program, and system should be underscored. Finally, an assessment system should establish a means through which assessor competence in conducting the selected assessment tools can be demonstrated. Typically, competence is either assumed, due to established qualifications or credentials, or assumed as the outcome of some targeted training. However, procedures for certifying that assessors can appropriately conduct assessments using the intended tools, with the target children, in the expected programs can be established.

CONSIDERATIONS AND CAVEATS:

- Best practice in early childhood assessment suggests that assessment of young children be conducted by either an individual with an existing relationship with the child, or by a previously unknown individual who takes time to establish a rapport with the child before the assessment begins.
- The complexity and administration demands of the selected assessment(s) inform the necessary skills and experiences for assessors. They also inform training and ongoing monitoring of administration required to ensure appropriate administration. For many direct assessments, assessors trained specifically on the administration of the assessment are typically used, although it is possible to train teachers on the administration of many measures, especially those that rely on observation, rather than direct assessment.
- In system-wide assessment programs, the qualifications, training, and role of the assessor should be common across the system. Variations here, including potential assessor effects (e.g., Waterman et. al 2012), threaten the validity of the results.
- Generally speaking, assessments completed by individuals close to the child are well suited for some purposes (e.g., informing instruction), but tend to be discouraged when used for other purposes (e.g., evaluation).

When and how often are assessments completed?

There are two time dimensions to consider in defining a kindergarten readiness assessment system—the frequency of assessment and when it occurs. To most effectively inform instruction, assessment should be ongoing to provide continuous feedback to the teacher on children's progress. To be used for program evaluation purposes, multiple assessments are necessary to allow for the measurement of growth during the program. As a result, decisions about the frequency of assessment are tied to the purpose of the assessment.

However, the use of readiness assessments for children once they are enrolled can provide important information about the children being served and areas of comparative need.

In considering the timing of kindergarten readiness in relation to date of entry, there is an important distinction to be made between kindergarten readiness assessment as a means of screening children for placement in kindergarten and kindergarten readiness assessment as a means of determining the general level of abilities of entering kindergartners. In general, using readiness assessments to place children into kindergarten classes has had limited success (e.g., Meisels 1987; La Paro & Pianta 2000), and using kindergarten readiness assessments as means of limiting access to kindergarten is viewed as inappropriate practice (NAEYC & NAECS/SDE 2001). However, the use of readiness assessments for children once they are enrolled can provide important information about the children being served and areas of comparative need (e.g., Brown et al. 2007). By definition, kindergarten readiness assessments used to

screen children into kindergarten must occur prior to kindergarten entry (and they may result in denying access). Assessing the skill set of children about to enroll or enrolled in kindergarten may occur in the months prior to kindergarten entry, but most typically occurs once the kindergarten year has begun.

CONSIDERATIONS AND CAVEATS:

- Single point-in-time assessments can provide a snapshot of children's performance, but cannot be used to examine change in children over time.
- Single point-in-time snapshots may be used to examine successive cohorts of children entering kindergarten in a state or locality to estimate changes in the populations of children served over time.
- To describe change over time in children, whether the purpose is to inform instruction or to evaluate programs, assessments of the same children must be completed at multiple time points. Single point-in-time assessment data cannot provide developmental data on individual children.
- When using a single point in time to assess children once enrolled in kindergarten, the timing should allow for a brief period of adjustment to the classroom and school routines, but not be delayed so far into the school year that the results are driven not only by kindergarten readiness but also by children's response to instruction during the first part of the kindergarten year. To allow comparability across groups of children, assessments should be implemented at about the same time for all children to avoid variations in performance based in part on children's variations in exposure to instruction.
- The majority of young children have experiences outside of the home during the year prior to kindergarten and they are spread across a wide range of programs and settings. Because of this dispersion in the population of children, conducting kindergarten readiness assessments prior to the beginning of the kindergarten year can be logistically and practically daunting.

What children are assessed?

The NAEYC and NAECS/SDE position statement on curriculum, assessment, and instruction advocates for the use of ongoing assessment to inform instructional progress and guide instruction to be developmentally appropriate for each child (NAEYC & NAECS/SDE 2003). This system assumes that all children are engaged in the assessment process. However, that is not necessarily the case depending upon the purpose of assessment and how the results will be used. Assessments used to evaluate programs, for example, can utilize a sample of children from the targeted program to estimate effects. Reducing the number of children assessed reduces the overall time burden on programs, or maintains the overall time burden but allows for longer assessment times per child. An additional alternative for evaluation purposes is matrix sampling. Under this design, a larger assessment is broken into smaller pieces, and each child is administered a portion of the total assessment. The data can then be combined to estimate program effects. The

decision to assess all children, or a sample, affects both cost and potential uses of the assessment data.

CONSIDERATIONS AND CAVEATS:

- The intention to assess all children or a sample of children should be informed by the purpose of the assessment and how the data will be used. The total number of children to be assessed has implications for costs as well as time burden on children and programs.
- To use assessments to screen for developmental disabilities, or to guide instruction, it is necessary for all children to be assessed; however, for other purposes, sampling strategies may be used.
- If a sample of children is to be used for a statewide assessment program, all children should participate in local or other assessment programs to provide information on their progress to inform instruction.
- If a sample of children is to be used, the sample should be designed to allow for estimates for children of different groups and should be based upon a random draw. The sample should provide adequate statistical power to establish estimates with minimal sampling error.
- An alternative sample-based approach, called matrix sampling, breaks a larger assessment into smaller pieces, and each child is administered a portion of the total assessment (see, e.g., Childs & Jaciw 2003). The data can then be combined to estimate program effects. This approach is taken in the National Assessment of Education Progress (NAEP), the so-called Nation's Report Card.
- Considerations should also be given for identifiable groups of children. For example, accommodations may or may not be necessary for children with disabilities. Likewise, children who are dual language learners, and may or may not be assessed in multiple languages, would be an important group to identify. Being able to identify groups based upon known risk factors, such as income (typically denoted as eligibility for free or reduced-price lunch in K–12 systems, possibly by receipt of services through federally funded programs for early education) and race/ethnicity may also be necessary if the purpose of the assessment is to measure effectiveness.

This diversity in linguistic and cultural backgrounds presents a number of complex considerations for early childhood education, generally, and assessment in early childhood especially.

HOW ARE ASSESSMENTS IN MULTIPLE LANGUAGES MANAGED?

There is a well-documented trend in the U.S. population toward greater diversity, including greater presence of young children coming from homes where English is either secondary or not spoken at all. While many of these children are exposed to English through programs prior to school entry, many are not, or many are exposed but have very limited English skills. This diversity in linguistic and cultural backgrounds presents a number of complex considerations for early childhood education, generally, and

assessment in early childhood especially (e.g., Espinosa 2005). How these children are assessed is driven a great deal by the purpose of the assessment and how the data will be used. For example, a Spanish language version of a math assessment will be a more accurate indicator of a Spanish speaking child's math abilities than will an assessment of mathematics given in English. However, the assessment in English may be a better index in the child's progress toward showing mathematics skills in the language in which he or she is most likely to be assessed as an older child.

CONSIDERATIONS AND CAVEATS:

- The NAEYC position statement on assessment of young ELLs (2005) captures the research consensus that ELL children be assessed in their home language where possible, and assessed in English only if they have an appropriate level of proficiency to provide for reasonable estimates of skills when assessed in English.
- The language of assessment is tied to the purpose of the assessment. If the purpose is to assess children's understanding of concepts or underlying skills, then assessment in the child's primary language may be appropriate. If the purpose is to assess the child's progress in English, then assessment in English would be more important.
- Will the assessment be conducted in languages other than English? If so, are there appropriate versions of the same assessment, with documented psychometric properties that are sound for all languages of administration?
- How will data from multiple languages be pooled, reported, and analyzed?
- If assessing in multiple languages, how will children be screened for assessment in a non-English language? The Early Childhood Longitudinal Study, Birth Cohort (ECLS-B) (Najarian et al. 2010) study utilized a screen with a relatively low level of English proficiency required for children to be assessed in English, while the Early Childhood Longitudinal Study, Kindergarten Cohort of 1998 (ECLS-K) (Rock & Pollack 2002) utilized a higher level of demonstrated proficiency.
- If being assessed in a language other than English, the assessments should be conducted by assessors fluent in the language of assessment.
- If assessed in languages other than English, the language of assessment should be reported and the results of assessment interpreted as the child's skills or abilities as demonstrated in English. This avoids confusion between demonstration of skill in a second language for dual language learners and the potential for a different level of skill when assessed in another language.

UTILIZATION

The expected uses of data resulting from an assessment are tied to the purpose of the assessment, as noted earlier is discussing the section of assessment tools. When discussing utilization, the focus is on the outcome of the selected assessments—the data that they produce. Assessments of young children yield data tied to the assessment. These data are then used by teachers, caregivers, parents, program developers, policy makers, and others in a range of ways. When data result from an assessment intended to guide instruction, immediate access to and use of data are important in guiding teachers' classroom activities, so timeliness is tied to the usefulness of the data. When data from assessments are intended to serve as means of monitoring programs or estimating impact, additional steps may need to be taken to integrate data, ensure their quality, and generate analytic models to measure effectiveness. For these uses, more time is needed to ensure appropriate preparation for and use of the data, but the relevance of the data for making programmatic or policy decisions will decline over time. Beyond serving the immediate purpose(s) of assessment, data emanating from assessment systems can be used more systematically. The integration of assessment data with other data within a larger data system can support federal and state program development and inform policies. Such datasets may also be critical sources of new knowledge within the research community. Finally, once assessment data have been generated, a number of considerations concerning data confidentially and security arise that warrant discussion.

When discussing utilization, the focus is on the outcome of the selected assessments—the data that they produce.

TIMELINESS OF USE

Assessments are useful tools in so far as they are able to provide useable information, regardless of the purpose of assessment, in a timely way. However, stakeholders must balance the desire to use data quickly to inform educational practices and the need to assure appropriate quality in completing scoring and interpreting the data to provide feedback on program effectiveness. For teachers conducting authentic assessment to inform their instruction, the connection time between assessment and use of the result may be minutes. Other assessments strategies, including those that involve large-scale standardized assessments, or the completion of a large number of direct child assessments that require analyses before scores may be reported, often require data processing and analyses, potentially causing a delay of weeks or months between assessment and reporting. Whether assessments yield data very quickly or require substantial analyses before reporting, assessment data should be captured through secure data systems and processed through data quality procedures. These processes ensure that data are secure and error-free, but also create a lag

Stakeholders must balance the desire to use data quickly to inform educational practices and the need to assure appropriate quality in completing scoring and interpreting the data to provide feedback on program effectiveness.

between when assessments are administered and when their results may be known. As a result, it is important to develop realistic expectations for how quickly data from assessments can be used.

CONSIDERATIONS AND CAVEATS:

- The desire for rapid turnaround between assessment and data must be balanced with time to complete necessary data quality, and potentially data analytic, processes.
- Assessments used to inform instruction or screen for potential risk should yield results as quickly as possible.
- When data reporting lags behind the date of assessment, the discrepancy should be shared; data from a fall 2009 kindergarten readiness assessment that are released the following year should clearly be identified as being based upon the kindergartners from the previous fall.

Use of assessment data to ensure developmentally appropriate early education

The first two purposes for assessment articulated by the Education Goals Panel focus on individual children—using assessment to guide instruction and using instruction to identify children who may be in need of special services. Assessments to inform instruction provide data that can guide teachers toward instructional approaches to reach children's learning needs. Using assessments this way provides teaches with valuable feedback on their instruction and allows them to tailor it to more effectively support children's development. Likewise, assessments used to screen for the possible need for special services should lead directly to a response, in this case, more thorough assessment to move toward diagnostics. Assessments used to inform instruction or screen for potential weaknesses are invaluable for improving early education for children, but are not well suited for other purposes.

CONSIDERATIONS AND CAVEATS:

- Ongoing assessment, especially when it is conducted within the context of typically occurring activity (e.g., authentic assessment) provides information to teachers about children's current competencies that can be used to inform instruction.
- Assessments designed to inform instruction and to identify potential need for additional services should be administered to all children so that all children may benefit from them.
- Data obtained from assessments used for screening purposes should be linked to a systematic sequence of events when a child is identified to be at risk. This system should include additional diagnostic assessment and the provision of services to meet any identified special needs (Meisels & Atkins-Burnett 2005).
- Assessments should not be used to determine a child's eligibility for enrollment in kindergarten or continued receipt of kindergarten instruction.

Use of assessment data for evaluation and accountability purposes

The third and fourth purposes for assessment outlined by the National Education Goals Panel focus on the use of assessments to evaluate program effectiveness and for program accountability. While there is agreement that early childhood programs require evaluation, there is growing concern about the appropriateness of some early childhood assessment measures when used to evaluate programs and in the context of high-stakes assessment. As the National Academy's report Eager to Learn (Bowman, Donovan, & Burns 2001) noted, "uses of assessment data for purposes external to the classroom, rather than to improve educational practice directly, place a particularly heavy burden both on the assessment instruments and on the responsible adults." One consequence of this is that assessments used for other purposes, such as to inform instruction, are adapted, modified, or even misused for evaluation and accountability purposes. Snow and Van Hemel (2008, 343–344) underscore this in concluding that "there are not many tools designed for large scale program evaluation, so tools designed for other purposes often are adapted (e.g., shortened or administered differently) out of necessity, without sufficiently investigating the validity of the adapted tools in their new form and for their new purpose."

CONSIDERATIONS AND CAVEATS:

- While there is agreement that early childhood programs should require evaluation, there is growing concern about the appropriateness of some early childhood assessments as a means of evaluating programs and in the context of high-stakes assessment. The concern is especially with the use of measures intended to guide instruction or to screen for developmental problems used instead for accountability purposes.
- Because of the importance placed upon the results of studies examining program effectiveness, only assessments designed to yield such data and those with the greatest psychometric strength should be employed.
- Assessment results are only one data element in analyses of program effectiveness and on their own do little to clarify the relative roles of specific programs, prior life events, and other factors that correlate with school readiness.
- Assessment data used for evaluation or high stakes purposes do not need to be collected from all children. Assessing a sample of children can provide the data necessary for such analyses. In general, sampling costs less, and also requires less time, for children (especially when matrix sampling is used) than an assessment conducted with all children.

Use of assessment data in analyses within comprehensive data systems

As noted previously, the decisions about assessments and their implementation are important in ensuring that best practices in early childhood assessment are maintained. Ultimately, though, it is the results of these assessments (that is, the resulting scores) that will become part of data systems envisioned under ELC.

While the construction of complex, integrated data systems holds the promise for a wide range of positive outcomes for children, families, and programs, there is also the potential for harm. Once data systems are constructed, the temptation to analyze the data may precede the thoughtful, deliberate planning that is essential to ensure valid findings. Depth of data does not lead to depth of analytic capability.

Depth of data does not lead to depth of analytic capability.

CONSIDERATIONS AND CAVEATS:

- In using data from kindergarten readiness assessments, it is important to consider the children's age at assessment, as well as accounting for variations in how much time has elapsed since the beginning of the school year; both age at assessment and duration of school prior to assessment can result in differences in scores (see, e.g., Denton Flanagan & McPhee 2009).
- Assessment data captured in a comprehensive data system allow for a number of critical analyses, including the tracking of trends in the kindergarten population over time.
- Kindergarten readiness assessment data should be analyzed in ways consistent with the purposes of the assessment and the psychometric strength of the assessment. Assessments used to inform instruction may yield assessment data not appropriate for evaluative purposes.
- Within larger data systems, the wealth of data leads to the expectation that important differences may be explained (e.g., the kindergarten entry achievement gap). However, in the absence of experimental manipulations, or sufficient data to model pre-existing differences among children, families, and programs (i.e., selection effects), the analyses that can be conducted yield correlational findings only.
- Readiness assessments do not ensure that data systems have captured the breadth of information about the child's prior developmental and educational experiences necessary to support analysis of readiness data as measures of the effectiveness of earlier experiences.
- Readiness assessments provide only one measure of a program's effectiveness that should be combined with data about the quality of the program, as well as non-program factors.

Need for data confidentiality and security

Data produced by child assessments, as well as other data that may be included in a comprehensive data system, or even a more modest, program-level data system, should be treated as confidential and stored according to best practices for ensuring confidentiality and data security.[8] Doing so requires activity before, during, and after assessments have been completed. Before assessment data are collected, data systems and processes should be in place to capture data as necessary. Parents should be informed of assessment plans, and any necessary consent obtained. In some situations, parents may also need to provide consent

for data collected locally to be shared into larger data systems. Throughout assessment and data collection, the physical and electronic storage of assessment data should be secured. In some cases this may also require removal of personally identifying information from assessment records to ensure confidentiality. As the data system is built, and especially if data are to be integrated into larger systems, additional steps may be necessary to ensure data security and confidentiality. If not already done, personal information may be stripped from data records in this process.

CONSIDERATIONS AND CAVEATS:

- Conducting assessments, especially as a part of a program to inform instruction, may not require parental consent, but consent may be necessary for conducting other assessments and/or for sharing the results of assessments (data).
- Federal and/or state programs may have requirements for consent and data protection procedures, while individual programs may need to develop such guidelines.
- Data from assessments used to evaluate programs can be made confidential by removal of personally identifying information without affecting the data usefulness in evaluation. However, data that are intended for use over time (i.e., longitudinal) must have some means of linking a specific child's data together over time, even if that information is not personally identifying.
- If data are to be stored and maintained for long periods, it may become necessary to obtain consent from the children for continued use of their data once they reach the age of legal consent.
- The process of obtaining consent provides an important opportunity for programs to engage families.

SUMMARY AND CONCLUSION

Early childhood assessment systems, properly developed and implemented, can contribute greatly to the success of early childhood programs. Systems that effectively screen for follow-up children at risk for developmental delays can identify young and very young children for intervention services. Systems that inform a teacher's instruction better allow for targeted instruction and support to further children's learning and development. Systems that provide a portrait of skills children have as they enter public school systems can inform curriculum decisions. And assessments that can provide evidence of growth tied to participation in programs can guide implementation and policy decisions. Effective early childhood assessment systems exist within a larger early childhood system that provides programs to young children and supports teachers' professional growth. In designing early childhood assessment systems, fundamental decisions made about instrument selection, administration, and data utilization are interconnected, and decisions made about one aspect of an

assessment system can drive other options. This paper is intended to provide background for state and local policy makers, systems' builders, and program developers and implementers as assessment systems for young children are contemplated. An intentionally designed system for assessing young children is a necessary component of effective early childhood programs.

REFERENCES

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

Bagnato, S.J., J.T. Neisworth, & K. Pretti-Frontczak. 2010. *LINKing Authentic Assessment & Early Childhood Intervention: Best Measures for Best Practices.* 2nd ed. Baltimore, MD: Brookes.

Begeny, J.C., & H. Buchanan. 2010. "Teachers' Judgment of Students' Early Literacy Skills Measured by the Early Literacy Skills Assessment: Comparisons of Teachers with and without Assessment Administration Experience. *Psychology in the Schools,* 47: 859–868.

Berry, D.J., L.J. Bridges, & M.J. Zaslow. 2004. *Early Childhood Measures Profiles.* Washington, DC: Child Trends. http://www.childtrends.org/Files/Child_Trends-2004_09_01_FR_ECMeasures.pdf.

Boller, K., S. Atkins-Burnett, L.M. Malone, G.P. Baxter, & J. West. 2010. *Compendium of Student, Teacher, and Classroom Measures Used in NCEE Evaluations of Educational Interventions. Volume I: Measures Selection Approaches and Compendium Development Methods* (NCC 2010-4012). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education. http://ies.ed.gov/ncee/pubs/20104012/index.asp.

Bowman, B.T., M.S. Donovan, & M.S. Burns, eds. 2001. *Eager to Learn: Educating Our Preschoolers.* Washington, DC: National Academy Press.

Brown, G., C. Scott-Little, L. Amwake, & L. Wynn. 2007. *A Review of Methods and Instruments in State and Local School Readiness Evaluations* (Issue & Answers Report, REL 2007-004). Washington, DC: U.S. Department of Education. Retrieved September 8, 2011, from http://ies.ed.gov/ncee/edlabs.

Childs, R.A., & A.P. Jaciw. 2003. "Matrix Sampling of Items in Large-Scale Assessments." *Practical Assessment, Research & Evaluation* 8 (16). Retrieved September 8, 2011, from http://PAREonline.net/getvn.asp?v=8&n=16

Daily, S., M. Burkhauser, & T. Halle. 2010. "A Review of School Readiness Practices in the States: Early Learning Guidelines and Assessments." *Early Childhood Highlights* 1 (3). Available at http://www.childtrends.org.

DEC (Division for Early Childhood). 2007. *Promoting Positive Outcomes for Children with Disabilities: Recommendations for Curriculum, Assessment, and Program Evaluation.* Missoula, MT: Author.

Denton Flanagan, K., & C. McPhee. 2009. *The Children Born in 2001 at Kindergarten Entry: First Findings From the Kindergarten Data Collections of the Early Childhood Longitudinal Study, Birth Cohort (ECLS-B)* (NCES 2010-005). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC: U.S. Department of Education. http://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2010005

Diamond, A. 2010. "The Evidence Base for Improving School Outcomes by Addressing the Whole Child and by Addressing Skills and Attitudes, Not Just Content." *Early Education & Development* 21: 780–793.

Espinosa, L.M. 2005. "Curriculum and Assessment Considerations for Young Children from Culturally, Linguistically, and Economically Diverse Backgrounds." *Psychology in the Schools* 42: 837–853.

Guhn, M., M. Janus, & C. Hertzman. 2007. "The Early Development Instrument: Translating School Readiness Assessment into Community Actions and Policy Planning." *Early Education and Development* 18 (3): 369–374.

Halle, T., M. Zaslow, J. Wessel, S. Moodie, & K. Darling-Churchill. 2011. *Understanding and Choosing Assessments and Developmental Screeners for Young Children Ages 3–5: Profiles of Selected Measures.* Washington, DC: Office of Planning, Research and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services.

Janus, M., S. Brinkman, E. Duke, C. Hertzman, R. Santos, M. Sayers, & J. Schroeder. 2007. *The Early Development Instrument: A Population-Based Measure for Communities. A Handbook on Development, Properties, and Use.* Hamilton, Ontario: Offord Centre for Child Studies.

Kowalski, K., R.D. Brown, & K. Pretti-Frontczak. 2005. "The Effects of Using Formal Assessment of Preschool Teachers' Beliefs about the Importance of Various Developmental Skills and Abilities." *Contemporary Educational Psychology* 30: 23–42.

La Paro, K.M., & R.C. Pianta. 2000. "Predicting Children's Competence in the Early School Years: A Meta-Analytic Review." *Review of Educational Research* 70 (4): 443–484.

Martone, A., & S.G. Sireci. 2009. "Evaluating Alignment between Curriculum, Assessment, and Instruction." *Review of Educational Research* 79: 1332–1361.

Meisels, S.J. 1987. "Uses and Abuses of Developmental Screening and School Readiness Testing." *Young Children* 42 (2): 68–73.

Meisels, S.J. 2007. "Accountability in Early Childhood: No Easy Answers." In *School Readiness and the Transition to Kindergarten,* eds. R.C. Pianta, M.J. Cox, & K. Snow, 31–48. Baltimore: Paul H. Brookes.

Meisels, S.J., & S. Atkins-Burnett. 2004. "The Head Start National Reporting System: A Critique." *Young Children* 59 (1): 64-66.

Meisels, S.J., & S. Atkins-Burnett. 2005. *Developmental Screening in Early Childhood: A Guide.* 5th ed. Washington, DC: NAEYC.

NAEYC. 2005. *Screening and Assessment of Young English-Language Learners.* Washington, DC: Author. http://www.naeyc.org/files/naeyc/file/positions/WWSEnglishLanguageLearnersWeb.pdf.

NAEYC & NAECS/SDE (National Association of Early Childhood Specialists in State Departments of Education). 2001. *STILL Unacceptable Trends in Kindergarten Entry and Placement.* http://www.naeyc.org/files/naeyc/file/positions/Psunacc.pdf.

NAEYC & NAECS/SDE (National Association of Early Childhood Specialists in State Departments of Education). 2003. *Early Childhood Curriculum, Assessment, and Program Evaluation: Building an Effective, Accountable System in Programs for Children Birth through Age 8.* Washington, DC: NAEYC. http://www.naeyc.org/files/naeyc/file/positions/pscape.pdf.

Najarian, M., K. Snow, J. Lennon, & S. Kinsey. 2010. *Early Childhood Longitudinal Study, Birth Cohort (ECLS-B), Preschool–Kindergarten 2007 Psychometric Report* (NCES 2010-009). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC.: U.S. Department of Education. http://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2010009

Neisworth, J.T., & S.J. Bagnato. 2004. "The Mismeasure of Young Children: The Authentic Assessment Alternative. *Infants and Young Children* 17 (3): 198–212.

Roach, A.T., D. McGarth, C. Wixson, & D. Talapatra. 2010. "Aligning an Early Childhood Assessment to State Kindergarten Content Standards: Application of a Nationally Recognized Alignment Framework." *Educational Measurement: Issues and Practice* 29: 25–37.

Rock, D.A., & J.M. Pollack. 2002. *Early Childhood Longitudinal Study-Kindergarten Class of 1998-99 (ECLS-K), Psychometric Report for Kindergarten Through First Grade.* U.S. Department of Education, National Center for Education Statistics. Washington, DC: U.S. Department of Education. http://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=200205

Scott-Little, C., J. Lesko, J. Martella, & P. Millburn. 2007. "Early Learning Standards: Results from a National Survey to Document Trends in State-Level Policies and Practices." *Early Childhood Research & Practice* 9 (1). Retrieved September 15, 2010, from http://ecrp.uiuc.edu/v9n1/little.html.

Shepard, L.A. 1994. "The Challenges of Assessing Young Children Appropriately." *Phi Delta Kappan* 76: 206–212.

Shepard, L.A. 2000. "The Role of Assessment in a Learning Culture." *Educational Researcher* 29: 4–14.

Shepard, L.A., S.L. Kagan, & E.O. Wurtz. 1998. *Principles and Recommendations from Early Childhood Assessments.* Washington, DC: National Education Goals Panel.

Snow, K.L. 2006. "Measuring School Readiness: Conceptual and Practical Considerations." *Early Education and Development* 17: 7–41.

Snow, K.L. 2007. "Integrative Views of the Domains of Child Function." In *School Readiness and the Transition to Kindergarten in the Era of Accountability,* eds. R. Pianta, M. Cox, & K. Snow, 197–216. Baltimore: Brookes.

Snow, C.E., & S.B. Van Hemel. 2008. *Early Childhood Assessment: Why, What, and How?* National Research Council of the National Academies Report. Washington, DC: The National Academy Press.

Stedron, J., & A. Berger. 2010. NCSL Technical Report: State Approaches to School Readiness Assessment (updated June 2010). Denver, CO: National Conference of State Legislators. http://www.ncsl.org/documents/Educ/KindergartenAssessment.pdf.

Waterman, C., P.A. McDermott, J.W. Fantuzzo, & V.L. Gadsden. 2012. "The Matter of Assessor Variance in Early Childhood Education—Or Whose Score Is It Anyway?" *Early Childhood Research Quarterly* 27: 46–54.

ENDNOTES:

1 Program Standard 4: The program is informed by ongoing systematic, formal, and informal assessment approaches to provide information on children's learning and development. These assessments occur within the context of reciprocal communications with families and with sensitivity to the cultural contexts in which children develop. Assessment results are used to benefit children by informing sound decisions about children, teaching, and program improvement.

2 Candidates prepared in early childhood degree programs understand that child observation, documentation, and other forms of assessment are central to the practice of all early childhood professionals. They know about and understand the goals, benefits, and uses of assessment. They know about and use systematic observations, documentation, and other effective assessment strategies in a responsible way, in partnership with families and other professionals, to positively influence the development of every child.

3 Lack of funding is, of course, one of a number of factors that limit participation of eligible families. Lack of awareness, difficulty navigating procedures, and parental choice to not participate or use other types of programs are also factors in the limited reach of these federal programs.

4 In addition, states provide differing models of kindergarten programs, including transitional kindergarten and two-year kindergarten. Some states allow for kindergartners to be held back to repeat a year of kindergarten before progressing into first grade.

5 Throughout this paper, we use the terms "young children" and "early childhood" to mean children from birth to age 8, reflecting the NAEYC definition of early childhood. Most of the current impetus for assessment, however, is focused on children in preschool and older.

6 It is not the intent of this guidance to provide details about psychometrics. There are many assessment textbooks and other resources available to provide this depth of coverage.

7 For further discussion about assessing ELLs, see NAEYC (2005). For further discussion on assessing children with disabilities, see DEC (2007).

8 While most contemporary discussion about data security occurs within the context of electronic or digital systems, many of the same concerns arise when the data are captured and stored physically. This is especially likely when assessments are used to inform instruction (where data on responses may not be captured at all), or when assessment data are considered at a classroom or program level, where electronic systems are less likely to be present than in statewide or program-wide contexts.

naeyc®

National Association for the Education of Young Children
1313 L Street NW, Suite 500, Washington, DC 20005-4101
202-232-8777 800-424-2460
www.naeyc.org