nwea
Measuring What Matters

WHITE PAPER

How to navigate early childhood assessment

A research-based guide to inform assessment planning in the early grades

Cindy Jiban, PhD

TABLE OF CONTENTS

Research and professional best practices show us that early childhood is a place of tremendous opportunity, but it is also a place for care and consideration. Given the massive impact of appropriate, quality educational programs and interventions for children at these young ages, relying on the best sources of data to inform decisions is critical.

Even for professionals who make decisions about student assessment on a regular basis, the arena of early childhood assessment can be difficult to navigate. It is not enough to simply assess earlier content using the same approaches as those used in older grades, or to take decisions about tools and purposes that were made with older students in mind and extend them to younger children. Instead, professional standards and guidelines for early childhood assessment must begin with attention to the important reality that young children are continuously and rapidly developing—academically and across a wide range of other domains. The context that informs assessment decisions for early learners is qualitatively different from the context for older students.

The goal of this paper is to support leaders in planning and reviewing their assessment implementations in the early grades. This paper will help readers with the following:

+ Understand the 'big ideas' from early childhood thought leaders—what do they believe should guide assessment decisions for the youngest school-aged students (pre-kindergarten through third grade)?
+ Discover what the research shows to be effective in terms of assessment in the early grades
+ Come away with a clear sense of next steps to apply the research and best practices to your own assessment planning process

In preparing this paper, we reviewed key ideas from professional guidelines on early childhood assessment. To frame our analysis of these guidelines, we also addressed two topics: 1) background on assessment-related concerns in the early childhood field, and 2) evidence of the leverage that early education and intervention provide on later outcomes. In this context, we closely examined professional recommendations relating to assessment purpose and assessment method. Finally, taking these important considerations into account, we created three guides for best practices in the assessment planning process in the early grades.

REVIEWED GUIDELINES

We reviewed five seminal reports on professional guidelines for this paper. Together, these focus on children in pre-kindergarten through age eight. Key points from each document are included in the Appendix.

01 National Education Goals Panel (NEGP), 1998. In 1998, the NEGP convened a working group related to the following goal: "By the year 2000, all children in America will start school ready to learn." This working group produced a document entitled *Principles and Recommendations for Early Childhood Assessment*. These principles can be found in Appendix A.

02 National Association for the Education of Young Children (NAEYC) and National Association of Early Childhood Specialists in State Departments of Education (NAECS/SDE), 2003. In 2003, after the establishment of the No Child Left Behind law changed the landscape for educational assessment, NAEYC and the NAECS/SDE jointly drafted a position statement entitled *Early Childhood Curriculum, Assessment, and Program Evaluation*. Key assessment recommendations and indicators of effectiveness from this document can be found in Appendix B.

03 Division for Early Childhood (DEC), 2007. In 2007, the DEC of the Council for Exceptional Children developed a response to the 2003 position statement from NAEYC and NAECS/SDE. The DEC document highlights considerations for children with disabilities, but encompasses recommendations applicable to the broader community of which these children are members. The paper is called *Promoting Positive Outcomes for Children with Disabilities: Recommendations for Curriculum, Assessment, and Program Evaluation*. Key recommendation and critical attributes from this document can be found in Appendix C.

04 National Research Council (NRC), 2008. The NRC was commissioned to study important developmental outcomes for children through age five and to guide the appropriate assessment of these outcomes. In their 2008 book, *Early Childhood Assessment: Why, What, and How*, the NRC committee emphasized several essential principles. The NRC Guidelines on Purposes of Assessment, Instrument Selection and Implementation, and Systems can be found in Appendix D.

05 Council of Chief State School Officers' Early Childhood Education State Collaborative on Assessment and Student Standards (CCSSO Early Childhood Education SCASS), 2011. In the context of the federal Race to the Top—Early Learning Challenge state grant program calling for kindergarten readiness assessments as a critical element, CCSSO Early Childhood Education SCASS published a position paper. The key conditions for useful assessment asserted in the paper, Moving Forward with Kindergarten Readiness Assessment Efforts, can be found in Appendix E.

BIG IDEAS SHARED ACROSS DOCUMENTS

From these reviewed guidelines, three big ideas emerged as central concerns for all the authoring groups.

01 **PURPOSEFUL ASSESSMENT.** The design, use, and interpretation of assessments must be purpose driven. Too many negative outcomes derive from assessments of young children used for purposes for which they were not designed; the type of inferences made from assessment data must be determined in the context of each specific purpose.

02 **INSTRUCTIONALLY ALIGNED ASSESSMENT.** Assessments must be clearly and explicitly integrated into the overall system, including curriculum and instruction; material assessed must represent the valued outcomes on which instruction is focused. This includes reaching toward alignment to standards or curriculum, where these exist. For classroom-based assessments designed to inform instruction, this also encompasses alignment to the instructional calendar.

03 **BENEFICIAL ASSESSMENT.** Assessments of children must serve to optimize learning. Time and resources are taken away from instruction to assess—and historically, there has been some justification for the fear that assessment data may offer unintended negative consequences for some children (NRC, 2008). Assessments must demonstrate solid consequential validity: the consequence of the time and resources invested in the assessment should be demonstrably positive for the children assessed.

CHALLENGES IN EARLY CHILDHOOD ASSESSMENT

While these big ideas represent significant consensus, there is also a vein of debate running through the early childhood field. Some professionals voice concerns over the increasing emphasis on assessment of young children, often focusing particularly on standardized tests. To navigate these concerns with integrity, it is important to understand some of what is at root.

How do we account for developmental variability?

Professionals in early childhood education recognize deeply that typical, healthy children develop at different rates in different domains. It is an unusual child who is not early in developing in some domain and late in another—perhaps fine motor skills are developing more slowly, while language and social skills are zooming ahead. For some early childhood professionals, concerns arise about assigning younger children to static assessments designed to compare students to a proficiency norm, as has been common among state assessments for older children. Typically, the information produced by a static, proficiency-based test is weaker at greater distances from the proficiency mark. Given the greater intra- and inter-individual variability that younger children exhibit, some professionals are concerned that assessments may be used that offer low precision or information for children at lower and higher levels of achievement.

As the DEC (2007) notes, "Very young children learn and grow at remarkable and unpredictable rates that are unmatched during other age periods. Because of this, scores from assessments administered to very young children tend to be unstable" (p. 15). This has two repercussions for those with concerns. First, one-time snapshots are likely to be less meaningful for younger students, whose pace of growth exceeds that of older children. Second, professional judgment is a key factor in determining how ready each child is—particularly at and before kindergarten entry—for a certain approach to assessment.

The variability of young children's abilities relates to two key early childhood topics that carry significance for assessment: developmentally appropriate practice and opportunity to learn. NAEYC (2003) defines developmentally appropriate practice as pedagogy and care drawing from three sources of knowledge:

+ what we know about child development
+ what we know about each individual's interests, strengths, and weaknesses
+ what we know about the children's cultural and social context

The latter two kinds of knowledge relate to the need to be sensitive to a child's opportunity to learn. For instance, children who have never had adults read books with them in interactive ways have not yet had a chance to develop concepts about books and print. Good assessment practice needs to carefully attend to inferences made about children in cases when they are assessed on concepts they have not had the opportunity to learn (NRC, 2008, p. 357).

What gets measured?

Another concern in early childhood assessment stems from the possibility of mismatch between the narrow range of proficiencies that get measured and the breadth of proficiencies that children must develop—and programs must support—in early childhood. What is measured becomes what is taught, some fear; this might leave domains such as social and emotional development and creativity underemphasized in an assessment-driven atmosphere.

A group of early childhood professionals voiced this concern as the Common Core Standards in mathematics and literacy were drafted (Alliance for Childhood, 2010), and many continue to work toward expanding conversations to include other domains. To early childhood professionals, domains such as social development are central (NEGP, 1998; NAEYC & NAECS/SDE, 2003; DEC, 2007; NRC, 2008; CCSSO Early Childhood Education SCASS, 2011). Failure to assess these domains introduces a risk of failing to attend to them in instructional settings.

How should we assess?

Another concern is over the methods of assessment used. The Alliance for Childhood group has expressed concern that inappropriate and unreliable standardized tests might be used. Early childhood has some history of multi-method assessment, rich in indirect tools, such as interviews and tools that don't feel like assessments, like classroom observations. Seen from this perspective, the idea of "testing" may suggest to some a replacement of rich, multi-method assessment with a single tool that asks students to set aside their natural behavior or curiosity to answer a set of questions devoid of context.

However, issues with richer observational tools have also arisen. Assessment of kindergarteners has become a focus since the redoubling of federal attention under two grant programs, the Race to the Top—Early Learning Challenge grants (2011) and the Enhanced Assessment Grants (2013). Pushback has been significant, in some instances: these more comprehensive, authentic, observational kindergarten assessment approaches have proven time-intensive, and educators have raised concerns about their reliability and their efficiency (Gewertz, 2014; Flannery, 2015).

What purposes are appropriate?

A key consideration for any decision maker is the notion that the use and interpretation of assessments can have both positive and negative effects, both intended and unintended. No professional with integrity wants to see assessment data result in some children losing access to good instructional programming, for instance. The CCSSO Early Childhood SCASS particularly calls out a high-stakes version of the kindergarten readiness or school readiness test as inappropriate, noting that kids start kindergarten "at different ages and with different levels of knowledge and skills—even when they are developing quite typically" (p. 6). As the NRC committee on early childhood assessment (2008) notes, "Using readiness tests to make recommendations about children's access to kindergarten is especially troublesome because many of the children recommended for delayed entry are the ones who would most benefit from participation in an educational program" (p. 31).

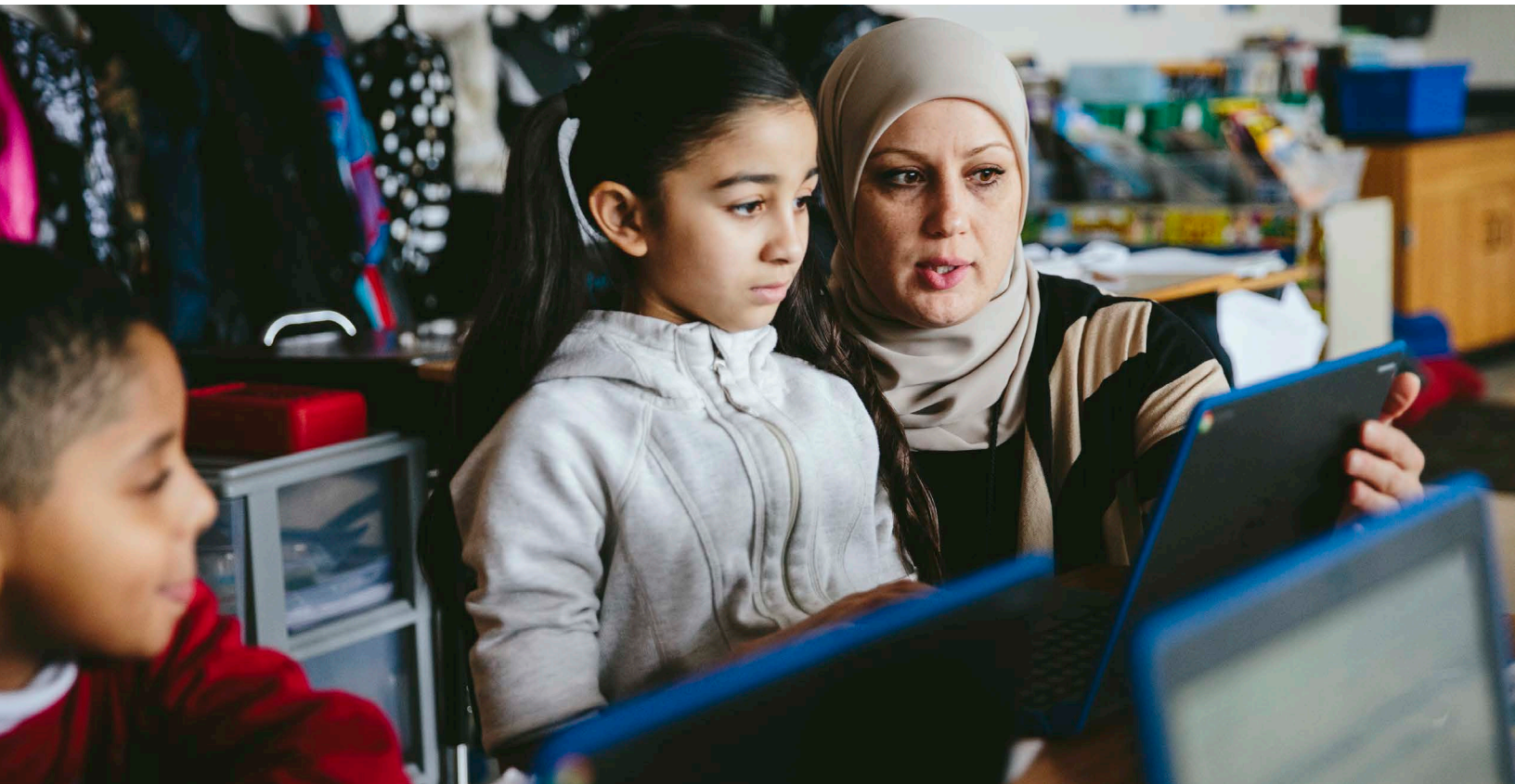
THE CASE FOR EVIDENCE-BASED EARLY INTERVENTION

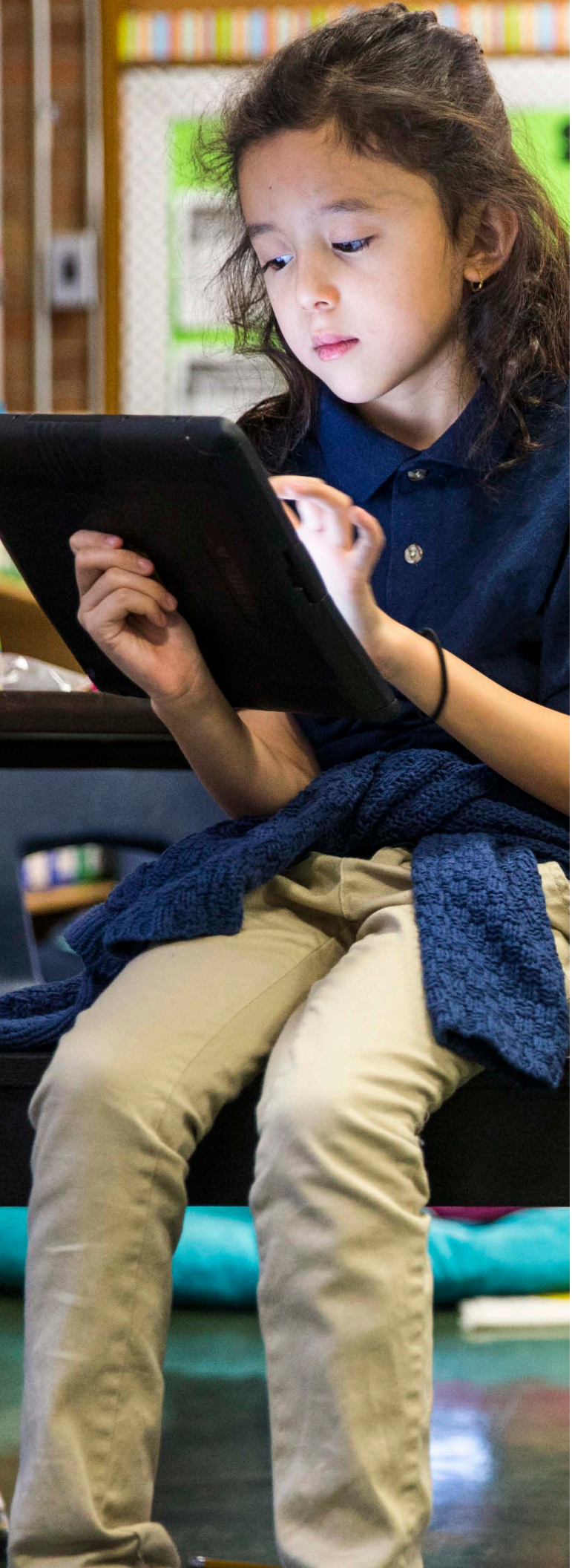
Why is it important to navigate these various concerns and arrive at high-quality assessment in early childhood? The defensible answer is because good data drives higher quality outcomes for children. To allocate limited resources with care, we must know who needs what, as well as which efforts succeed in meeting each child's needs. When resources go to sound, effective prevention and intervention efforts, early education offers leverage that is massive in comparison to efforts in the older grades.

The early childhood education field has a high degree of consensus around calls for universal access to quality preschool programming; educators also agree that learning in the primary grades is critically important. In reading and mathematics, in particular, the effectiveness of early intervention for preventing future difficulties is well supported in the literature (NRC, 1998; NAEYC & National Council of Teachers of Mathematics, 2002).

Early childhood educators can point to several US studies supporting the economic wisdom of investments in early intervention. A 2005 RAND study found that investments in early intervention programs offer a return to society from $1.80 to as much as $17.07 for every dollar spent. The topic gained traction in the summer of 2010, when the New York Times ran an article stating that economists "estimate that a standout kindergarten teacher is worth about $320,000 a year" based on the "additional money that a full class of students can expect to earn over their careers" because of the extra growth that teacher caused in kindergarten (Leonhardt, 2010).

Longitudinal studies of high-quality pre-kindergarten programs reveal that participating students benefit in a multitude of ways: increased high school graduation rates and decreased rates of both special education placement and crime or delinquency (Chicago Longitudinal Study); improved performance on standardized tests in later schooling and decreased chances of grade retention (Yale University Child Study Center); lower rates of teen pregnancy (The Carolina Abecedarian Project); and higher rates of employment and higher wages as adults (The High/Scope Perry Preschool Project) (Pre-K Now, 2010).





The case for early intervention in elementary schools, beginning in kindergarten, is clearest in the literature on how reading skills develop—or don't. When children enter school with poor pre-literacy skills, they are at a disadvantage; starting with lower initial skills strongly tends to lead to slower rates of growth in reading (Stanovich, 1986; Fielding, Kerr, & Rosier, 2007). As Juel (1988) explains, better readers read more words, increasing both their decoding and their vocabulary. Moreover, an early learner with broader deficits in oral language is at risk for poor growth in reading comprehension, even years later (Foorman et al., 2015).

As weak readers establish slower rates of reading, secondary problems often begin to emerge. Steele (2004) points to evidence of "frustration, anxiety, behavior problems, greater academic deficiencies, and subsequent motivation problems" developing for weaker readers. Perhaps in part because of these compounding issues, academic problems not identified prior to third grade are extremely resistant even to highly intense remedial efforts (Torgesen et al., 2001). Good, Simmons, and Smith (1998), members of the team creating DIBELS (Dynamic Indicators of Basic Early Literacy Skills), conclude the following:

> "[B]y the end of first grade and beginning of second grade, students on low developmental reading trajectories face nearly insurmountable obstacles to catching up with their peers. The answer lies in the early identification of children with deficits in crucial early literacy skills and enhancing their acquisition of those skills."

In mathematics, research evidence supports a similar "Matthew Effect," wherein the students who enter with weaknesses experience slower growth and gaps widen (Morgan, Farkas, & Wu, 2011). Researchers have focused on a critical early building block, number sense, which differentiates students at risk in terms of mathematical achievement as early as kindergarten (Jordan, Glutting, & Ramineni, 2010). When students begin school with a weak sense of how number and quantity work, their growth tends to be slowed across a breadth of mathematical topics. Fortunately, early screening for mathematical risk is becoming more and more accurate (Gersten et al., 2012). And, interventions for kindergarten and first-grade students at risk are proving increasingly successful at remediating skill deficits (e.g., Fuchs, Fuchs, & Compton, 2012).

When assessment data can work to match students at risk with effective early interventions, action is imperative. Mathematics achievement is increasingly important in our society. Low mathematical achievement in school decreases a student's chances of post-secondary educational opportunity

and increases risk for lifelong social and economic difficulties (Rivera-Batiz, 1992; National Mathematics Advisory Panel, 2008).

Two critical social concerns drive much of the push for the early identification of academic risk: 1) prevention of academic disabilities and 2) closure of academic achievement gaps across racial and socioeconomic groups. Given good, child-level data on the foundational proficiencies that are known to strongly predict future achievement in mathematics or literacy, disabilities can be prevented (Fletcher et al., 2007; Steele, 2004). A preventive framework that uses early and ongoing assessment to drive intervention can substantially reduce the number of students with learning disabilities (Gibbons, 2008).

Groups focused on achievement gaps have increasingly determined what the Association for Supervision and Curriculum Development (ASCD) stated in their 2006 infobrief: "Early intervention is the most cost-effective approach to closing the achievement gap" (p. 3). Perez-Johnson and Maynard (2007) note about their own research efforts: "Our focus on the period of early childhood stems from two critical research-based observations. First, early childhood is when achievement gaps first emerge. Second, early childhood represents an optimal period for intervention, because gaps compound and become more costly and difficult to address as time passes by" (p. 588).

CHOOSING ASSESSMENTS FOR EARLY LEARNING AND INTERVENTION

The remainder of this document takes the view that there is high potential value in reliable and valid data on student proficiencies, including in literacy and mathematics, and that direct assessments of students in pre-kindergarten through third grade have an appropriate role in generating such data. The five sets of professional guidelines mentioned earlier, all of which also take this view, are referenced here in greater detail for their guidance on assessment purposes and methods.

Purposes of assessment

Recommendations from all the major documents reviewed include an emphasis on purpose-driven assessment. Purposes should be explicit and public (NRC, 2008), specific (NAEYC/SDE, 2003; CCSSO Early Childhood Education SCASS, 2011), and beneficial (NEGP, 1998; NAEYC/NAECS/SDE, 2003; DEC, 2007; & NRC, 2008). All groups note that technical adequacy demands necessarily vary across purposes. The CCSSO Early Childhood Education SCASS particularly emphasizes that an assessment appropriate to a purpose, like classroom instructional planning, should not be enlisted in a high-stakes decision about program evaluation or comparing students across schools without a review of the tool particularly for this new purpose.

Broadly, the major documents agree that assessment purposes cluster around three main objectives: eligibility determination, instructional planning, and evaluation. The NEGP (1998, p. 7) offers these four categories:

+ assessments to support learning
+ assessments for identification of special needs
+ assessments for program evaluation and monitoring trends
+ assessments for high-stakes accountability

The set is further elaborated into seven specific purposes for children with disabilities by the DEC (2007). Most of these pertain to all students, as well:

01 screening

02 diagnosis (or identification) of delay or disability

03 eligibility determination for early intervention or special education services

04 instructional program planning/intervention assessment

05 placement

06 progress monitoring

07 program evaluation

Decisions for classroom instruction

Within the purposes which most closely pertain to classroom instruction (1, 4, and 6 above), three different questions about students are important. Screening addresses a question highly pertinent to schools using a Response to Intervention (RTI) model: Which students are at risk of poor outcomes in specific areas and should be offered more intensity of instruction?

A variety of test and non-test assessment information addresses another question for instructional or intervention planning: What do I still need to teach to my students or to a particular student? Finally, when instruction or interventions are underway, a question about effectiveness emerges: How much student progress is occurring?

The NRC committee on early childhood assessment (2008) describes it this way: "In addition to using assessment information to establish a descriptive picture of children's strengths and needs and to plan for instruction…teachers…need to collect ongoing assessment information to track their learning over time" (p. 32).

Methods of assessment

In early childhood, debate over assessment methods is strong. Nonetheless, from the five major sets of guidelines consulted here, consensus is evident around two main themes. First, assessment in early childhood should employ a variety of methods. Second, methods need to reach toward authentic assessment.

Multiple methods

To begin considering the principle of using multiple methods, we need to first know what methods make up the range of approaches used in early childhood assessment.

- NAEYC and NAECS/SDE (2003) list five methods in their definition of assessment: "Assessment: A systematic procedure for obtaining information from observation, interviews, portfolios, projects, tests, and other sources that can be used to make judgments about children's characteristics" (p. 27).
- The DEC (2007) offers the following list: "Potential assessment tools include: (1) record review/developmental history, (2) interviews, (3) observations, (4) checklists/rating scales, (5) portfolios, and (6) tests" (p. 13).
- The NRC committee on early mathematics learning recommendations (2009) offers three broad categories of assessment methods useful for formative purposes: observations, tasks, and interviews. Tasks include those on tests.

All the major sets of guidelines call for reliance upon a variety of assessment methods. In part, this point relates to the need for matching assessment purpose to assessment tool; different purposes will require different approaches. More particularly, good assessment will produce evidence about both what children can do and how children think about concepts. Both behaviors and cognitive explanations are valuable sources for the purpose of generating a comprehensive picture (NRC, 2009).

Tasks, including tests, tend to give evidence of what children can do, particularly when those tests are flexible or adaptive enough to reach toward a student's particular abilities. When insufficiently flexible, a test may result only in evidence of what a child cannot do; adaptive tests may offer greater reach toward showing where a child can succeed. Scaffolded tasks reach beyond what the child can do independently to show what the child is ready to do. Observations can also show what children can do, with some allowing children more latitude to initiate what kind of learning they are engaging with, with the assessment tool following their lead. Interviews allow teachers and students to "go beyond observation and tasks to probe the child's thinking" (NRC, 2009, p. 262). Assessment of young children should balance the task approach, observations, and interviews to robustly capture what students understand and can do (NEGP, 1998; NAEYC & NAECS/SDE, 2003; DEC, 2007; NRC, 2008; & NRC, 2009).

Authentic assessment

In addition to a call for multiple methods of assessment, the five major guideline documents also call for some version of "authenticity" in assessment methods. DEC (2007) explains that this encompasses tests, as well as other methods: "Tests may be norm-referenced, criterion-referenced, or curriculum-based; however, the most reliable outcomes for young children are generated when these tools are used within an authentic assessment model" (p. 13).

NAEYC and NAECS/SDE (2003) emphasize that assessment data should be gathered from "realistic settings and situations that reflect children's actual performance." Authentic assessment includes observations and tasks that occur in the context of regular play or activities, in settings typical to the child. They are "child-centered and interactive," resulting in more easily generalized information about "the child's ability to interact with the everyday environment." Authentic assessments capture "a large number of behaviors across multiple domains…[and] allow the child multiple opportunities to demonstrate a behavior or skill in multiple settings with preferred and multiple partners, objects, and materials, resulting in a more valid estimate of developmental status" (DEC, 2007, p. 14).

When students' performance on standard tasks or protocols is complemented with observations about what a child can do in 'real world' environments or with particular people, the resulting rich data may offer educators more insight into each child's unique strengths and needs.

PUTTING IT ALL TOGETHER

Research and professional best practices show us that early childhood is a place of tremendous leverage, but it is also a place for care and consideration. Given the potentially massive impact of appropriate, quality educational programs and interventions for children at these ages, relying on the best sources of data to inform decisions is critical.

In the early grades, an ideal assessment plan makes use of tools appropriate to each purpose and to young children; it also draws from multiple and authentic methods. It is designed to inform explicit decisions about eligibility, instructional planning, and effectiveness within domains that align to the educational program's goals for instruction and intervention. Most importantly, all aspects of the plan aim to clearly and effectively benefit the children assessed, while ensuring that no negative consequences are introduced. Good assessment planning and review provide a unique opportunity to do right by children and positively influence important outcomes in the near and long term.

BUILDING AN EARLY CHILDHOOD ASSESSMENT PLAN

The goal of planning a quality assessment solution for early learners can be met by applying the research and best practices reviewed above. To do so, a team of assessment and instructional leaders might work through assessment considerations in three related and cumulative steps. These steps are to consider alignment of assessment tools to:

01 domains of instruction or intervention

02 assessment purposes

03 assessment methods

To support assessment and instructional leaders in each of these three steps, sample tables are offered below as a framework for discussion and consideration.

STEP 1: DOMAINS: INSTRUCTION AND ASSESSMENT

In planning assessment implementations in the early grades (pre-kindergarten through third grade), it is important to begin with the educational program's goals for instruction and intervention as a starting point. Instruction and assessment should be aligned. Using the chart below, make explicit how what will be measured compares with what will be taught. Which domains are taught, but may not be assessed? For each area assessed, is there an opportunity to learn? Sometimes we assess in areas outside of learning—many programs screen for vision and hearing, for instance—but an intervention, such as communication with families, is the goal.

What is the domain of focus?	Will this be a goal for instruction or intervention?	Will this be assessed?
Example: Mathematics skill development	*Yes, both*	

STEP 2: ASSESSMENT PURPOSE: INFORMING EXPLICIT DECISIONS

In building a plan for assessment in each domain of interest, purpose is an important starting point. Purposes should translate clearly into decisions. In sorting out purposes for assessment, one might ask: What decisions should be informed by the resulting data? By pinning these down, it becomes possible to see whether and how each decision results in a clear benefit to the students being assessed.

We can think of assessment purposes as clustering broadly into decisions about eligibility, instructional planning, and effectiveness (evaluation). For each of these three categories, in the three tables below, list which decisions should be informed by data, which type of assessment is designed for this purpose, and which specific assessment tools you are considering. An example is offered in the domain of early literacy, drawing from a tiered intervention model like RTI.

Eligibility		
What data-driven decision do we need to make?	**What type of assessment can help inform this decision?**	**What are the candidate assessment tools we can use?**
Example: Which students may need intervention in early literacy?	*Screening*	

Instructional planning		
What data-driven decision do we need to make?	**What type of assessment can help inform this decision?**	**What are the candidate assessment tools we can use?**
Example: In which specific areas of literacy does the student need intervention?	*Instructional planning/ skills diagnostics*	

Effectiveness (evaluation)		
What data-driven decision do we need to make?	**What type of assessment can help inform this decision?**	**What are the candidate assessment tools we can use?**
Example: How much progress are students making in the intervention program?	*Progress monitoring*	

STEP 3: ASSESSMENT METHODS: MULTIPLE AND AUTHENTIC

Once the decisions that need to be informed by assessment data have been inventoried across all domains, the full list of candidate assessments can be reviewed. A review should certainly include a hard look at technical adequacy: Does a tool have the level of reliability, validity, and other properties required for informing a particular decision?

Another important review of candidate tools, though, should focus on the method of assessment itself. For instance, are all of the candidate tools tests, or do some enlist observations, interviews, or embedded tasks? Which make use of technology? Which provide evidence of what a student can do, or is ready to do, as compared with what a student cannot do? Which reach most toward authenticity by making use of realistic and everyday situations, incorporating an element of interaction or accommodating multiple behaviors across multiple settings and situations? It is important to note that authenticity does not simply go hand in hand with a particular method. Each method of assessment, whether an observational tool or a test, may incorporate more or fewer features of authenticity (such as use of feedback or of activities and materials that are like those used in the classroom or home setting).

For each of the candidate tools identified in Step 2, fill in the table below to paint a clear picture of the various assessment methods under consideration—as well as the degree to which the various tools incorporate authentic features.

What is the candidate assessment tool?	**What is the assessment method used?**	**Where/How is it administered?**	**What information/ evidence does it provide about students (i.e., can do, cannot do, ready to do)?**	**Which features of this tool reach toward authenticity?**
Example: Learning behaviors checklist	*Observation*	*Classroom; teacher completes checklist on paper*	*Can the student do the learning behaviors (e.g., following directions, asking questions)?*	*Takes place in student's everyday context, across multiple settings and situations, with multiple opportunities for evidence.*

REFERENCES

Alliance for Childhood. (2010). *Joint statement of early childhood health and education professionals on the Common Core Standards Initiative.* Retrieved October 28, 2010 from http://www.empoweredbyplay.org/2010/03/alliance-for-childhoods-joint-statement-of-early-childhood-health-and-education-professionals/

Association for Supervision and Curriculum Development. (2006). Closing the gap: Early childhood education. *Infobrief*, 45, 1-8.

Council of Chief State School Officers. (2011). Moving forward with kindergarten readiness assessment efforts: A position paper of the Early Childhood Education State Collaborative on Assessment and Student Standards. Washington, DC: *Council of Chief State School Officers*.

Division for Early Childhood. (2007). *Promoting positive outcomes for children with disabilities: Recommendations for curriculum, assessment, and program evaluation.* Missoula, MT: Author.

Fielding, L., Kerr, N., & Rosier, P. (2007). *Annual growth for all students, catch-up growth for those who are behind.* Kennewick, WA: The New Foundation Press, Inc.

Flannery, M. E. (2015, January 15). Kindergarten readiness tests wasting valuable teaching time. NEA Today. Retrieved from https://www.neatoday.org

Fletcher, J. M., Lyon, G. R., Fuchs, L. S., & Barnes, M. A. (2007). *Learning disabilities: From identification to intervention.* New York: Guildford Press.

Foorman, B.R., Herrara, S., Petscher, Y., Mitchell, A., & Truckenmiller, A. (2015). The structure of oral language and reading and their relation to comprehension in Kindergarten through Grade 2. Reading and Writing, 28, 655-681.

Fuchs, L. S., Fuchs, D., & Compton, D. L. (2012). The early prevention of mathematics difficulty: Its power and limitations. *Journal of Learning Disabilities*, 45(3), 257-269.

Gersten, R., Clarke, B., Jordan, N. C., Newman-Gonchar, R., Haymond, K., & Wilkins, C. (2012). Universal screening in mathematics for the primary grades: Beginnings of a research base. *Exceptional Children*, 78(4), 423-445.

Gewertz, C. (2014, October 7). Kindergarten-readiness assessments gain ground. *Education Week.* Retrieved from https://www.edweek.org

Good, R. H., Simmons, D. C., & Smith, S. B. (1998). Effective academic interventions in the United States: Evaluating and enhancing the acquisition of early reading skills. *School Psychology Review*, 27(1), 45-56.

Jordan, N. C., Glutting, J., & Ramineni, C. (2010). The importance of number sense to mathematics achievement in first and third grades. *Learning and Individual Differences*, 20, 82-88.

Juel, C. (1988). Learning to read and write: A longitudinal study of 54 children from first through fourth grades. *Journal of Educational Psychology*, 80, 437–447.

Leonhardt, D. (2010, July 27). The case for $320,000 kindergarten teachers. New York Times. Retrieved October 28, 2010 at http://www.nytimes.com/2010/07/28/business/economy/28leonhardt. html?_r=1

Morgan, P. L., Farkas, G., & Wu, Q. (2011). Kindergarten children's growth trajectories in reading and mathematics: Who falls increasingly behind? *Journal of Learning Disabilities*, 44(5), 472-488.

National Association for the Education of Young Children (NAEYC) and the Fred Rogers Center for Early Learning and Children's Media at Saint Vincent College (2012). *Technology and Interactive Media as Tools in Early Childhood Programs Serving Children from Birth through Age 8*. Retrieved May 2012 at http://www.naeyc.org/content/technology-and-young-children

National Association for the Education of Young Children & National Association of Early Childhood Specialists in State Departments of Education. (2003). *Early childhood curriculum, assessment, and program evaluation*. Washington, DC: National Association for the Education of Young Children.

National Education Goals Panel. (1998). *Principles and recommendations for early childhood assessments* (L. Shepard, S. L. Kagan, & E. Wurtz, Eds.). Washington, DC: Author.

National Research Council. (2009). *Mathematics learning in early childhood: Paths toward excellence and equity* (C. T. Cross, T. A. Woods, & H. Schweingruber, Eds.). Washington, DC: National Academies Press.

National Research Council. (2008). *Early childhood assessment: Why, what, and how* (C. E. Snow & S. B. Van Hemel, Eds.). Washington, DC: National Academies Press.

National Research Council. (1998). *Preventing reading difficulties in young children* (C.E. Snow, M.S. Burns, & P. Griffin, Eds.). Washington, DC: National Academy Press.

Pre-K Now (2010). *The benefits of high-quality pre-K.* Retrieved October 28, 2010 at http://www.preknow.org/advocate/factsheets/benefits.cfm

Perez-Johnson, I., & Maynard, R. (2007). The case for early, targeted interventions to prevent academic failure. *Peabody Journal of Education*, 82(4), 587-616.

RAND. (2005). *Proven benefits of early childhood interventions: Research brief.* Santa Monica, CA: Author.

Rivera-Batiz, F. L. (1992). Quantitative literacy and the likelihood of employment among young adults in the United States. *Journal of Human Resources*, 27, 313–328.

Steele, M. (2004). Making the case for early identification and intervention for young children at risk for learning disabilities. *Early Childhood Education Journal*, 32(2), 75-9.

Torgesen, J.K., Rashotte, C.A., & Alexander, A.W. (2001). Principles of fluency instruction in reading: Relationships with established empirical outcomes. In M. Wolf (Ed.), *Dyslexia, fluency, and the brain* (pp. 333–335). Timonium, MD: York Press.

Wurtzel, J., Marion, S., Perie, M., and Gong, B. (2009). The role of interim assessments in a comprehensive assessment system. In L. Pinkus (Ed.), *Meaningful measurement: The role of assessments in improving high school education in the Twenty-First Century* (pp. 77-94). Washington, DC: Alliance for Excellent Education.

APPENDIX

Early childhood assessment

APPENDIX A: NEGP ASSESSMENT PRINCIPLES
National Education Goals Panel (NEGP), 1998, pp. 5-6

- Assessment should bring about benefits for children. Gathering accurate information from young children is difficult and potentially stressful. Formal assessments may also be costly and take resources that could otherwise be spent directly on programs and services for young children. To warrant conducting assessments, there must be a clear benefit—either in direct services to the child or in improved quality.
- Assessments should be tailored to a specific purpose and should be reliable, valid, and fair for that purpose. Assessments designed for one purpose are not necessarily valid if used for other purposes. In the past, many of the abuses of testing with young children have occurred because of misuse. The recommendations in the sections that follow are tailored to specific assessment purposes.
- Assessment policies should be designed recognizing that reliability and validity of assessments increase with children's age. The younger the child, the more difficult it is to obtain reliable and valid assessment data. It is particularly difficult to assess children's cognitive abilities accurately before age 6. Because of problems with reliability and validity, some types of assessment should be postponed until children are older, while other types of assessment can be pursued, but only with necessary safeguards.
- Assessments should be age-appropriate in both content and the method of data collection. Assessments of young children should address the full range of early learning and development, including physical well-being and motor development; social and emotional development; approaches toward learning; language development; and cognition and general knowledge. Methods of assessment should recognize that children need familiar contexts in order to be able to demonstrate their abilities. Abstract paper-and-pencil tasks may make it especially difficult for young children to show what they know.
- Assessments should be linguistically appropriate, recognizing that to some extent all assessments are measures of language. Regardless of whether an assessment is intended to measure early reading skills, knowledge of color names, or learning potential, assessment results are easily confounded by language proficiency, especially for children who come from home backgrounds with limited exposure to English, for whom the assessment would essentially be an assessment of their English proficiency. Each child's first- and second-language development should be taken into account when determining appropriate assessment methods and in interpreting the meaning of assessment results.
- Parents should be a valued source of assessment information, as well as an audience for assessment results. Because of the fallibility of direct measures of young children, assessments should include multiple sources of evidence, especially reports from parents and teachers. Assessment results should be shared with parents as part of an ongoing process that involves parents in their child's education.

APPENDIX B: NAEYC AND NAECS/SDE RECOMMENDATION AND INDICATORS OF ASSESSMENT EFFECTIVENESS
National Association for the Education of Young Children (NAEYC) and National Association of Early Childhood Specialists in State Departments of Education (NAECS/SDE), 2003, pp.10-11

Key Recommendation: Make ethical, appropriate, valid, and reliable assessment a central part of all early childhood programs. To assess young children's strengths, progress, and needs, use assessment methods that are developmentally appropriate, culturally and linguistically responsive, tied to children's daily activities, supported by professional development, inclusive of families, and connected to specific, beneficial purposes: (1) making sound decisions about teaching and learning, (2) identifying significant concerns that may require focused intervention for individual children, and (3) helping programs improve their educational and developmental interventions.

Indicators of Effectiveness:

- Ethical principles guide assessment practices.
- Assessment instruments are used for their intended purposes.
- Assessments are appropriate for ages and other characteristics of children being assessed.
- Assessment instruments are in compliance with professional criteria for quality.
- What is assessed is developmentally and educationally significant.
- Assessment evidence is used to understand and improve learning.
- Assessment evidence is gathered from realistic settings and situations that reflect children's actual performance.
- Assessments use multiple sources of evidence gathered over time.
- Screening is always linked to follow-up.
- Use of individually administered, norm-referenced tests is limited.
- Staff and families are knowledgeable about assessment.

APPENDIX C: DEC RECOMMENDATION AND CRITICAL ATTRIBUTES FOR ASSESSMENT
Division for Early Childhood (DEC), 2007, pp. 10-15

Key Recommendation: Assessment is a shared experience between families and professionals in which information and ideas are exchanged to benefit a child's growth and development. Assessment practices should be integrated and individualized in order to: (a) answer the questions posed by the assessment team (including family members); (b) integrate the child's everyday routines, interests, materials, caregivers, and play partners within the assessment process; and (c) develop a system for shared partnerships with professionals and families for the communication and collection of ongoing information valuable for teaching and learning. Therefore, assessment teams should implement a child- and family-centered, team-based, and ecologically valid assessment process. This process should be designed to address each child's unique strengths and needs through authentic, developmentally appropriate, culturally and linguistically responsive, multidimensional assessment methods. The methods should be matched to the purpose for the assessment, linked to curriculum and intervention, and supported by professional development.

Critical Attributes:

- Assessment tools have utility and are used for specific purposes.
- Assessment tools are authentic.
- Assessment tools have good psychometric qualities.

APPENDIX D: NRC GUIDELINES ON PURPOSES OF ASSESSMENT AND ON INSTRUMENT SELECTION AND IMPLEMENTATION

National Research Council (NRC), 2008, pp. 5-9

Guidelines on Purposes of Assessment

- (P-1) Public and private entities undertaking the assessment of young children should make the purposes of assessment explicit and public.
- (P-2) The assessment strategy—which assessments to use, how often to administer them, how long they should be, how the domain of items or children or programs should be sampled—should match the stated purpose and require the minimum amount of time to obtain valid results for that purpose. Even assessments that do not directly involve children, such as classroom observations, teacher rating forms, and collection of work products, impose a burden on adults and will require advance planning for using the information.
- (P-3) Those charged with selecting assessments need to weigh options carefully, considering the appropriateness of candidate assessments for the desired purpose and for use with all the subgroups of children to be included. Although the same measure may be used for more than one purpose, prior consideration of all potential purposes is essential, as is careful analysis of the actual content of the assessment instrument. Direct examination of the assessment items is important because the title of a measure does not always reflect the content.

Guidelines on Instrument Selection and Implementation

- (I-1) Selection of a tool or instrument should always include careful attention to its psychometric properties.
  - A. Assessment tools should be chosen that have been shown to have acceptable levels of validity and reliability evidence for the purposes for which they will be used and the populations that will be assessed.
  - B. Those charged with implementing assessment systems need to make sure that procedures are in place to examine validity data as part of instrument selection and then to examine the data being produced with the instrument to ensure that the scores being generated are valid for the purposes for which they are being used.
  - C. Test developers and others need to collect and make available evidence about the validity of inferences for language and cultural minority groups and for children with disabilities.
  - D. Program directors, policy makers, and others who select instruments for assessments should receive instruction in how to select and use assessment instruments.
- (I-2) Assessments should not be given without clear plans for follow-up steps that use the information productively and appropriately.
- (I-3) When assessments are carried out, primary caregivers should be informed in advance about their purposes and focus. When assessments are for screening purposes, primary caregivers should be informed promptly about the results, in particular whether they indicate a need for further diagnostic assessment.
- (I-4) Pediatricians, primary medical caregivers, and other qualified personnel should screen for maternal or family factors that might impact child outcomes—child abuse risk, maternal depression, and other factors known to relate to later outcomes.
- (I-5) Screening assessment should be done only when the available instruments are informative and have good predictive validity.
- (I-6) Assessors, teachers, and program administrators should be able to articulate the purpose of assessments to parents and others.
- (I-7) Assessors should be trained to meet a clearly specified level of expertise in administering assessments, should be monitored systematically, and should be reevaluated occasionally. Teachers or other program staff may administer assessments if they are carefully supervised and if reliability checks and monitoring are in place to ensure adherence to approved procedures.
- (I-8) States or other groups selecting high-stakes assessments should leave an audit trail—a public record of the decision making that was part of the design and development of the assessment system. These decisions would include why the assessment data are being collected, why a particular set of outcomes was selected for assessment, why the particular tools were selected, how the results will be reported and to whom, as well as how the assessors were trained and the assessment process was monitored.
- (I-9) For large-scale assessment systems, decisions regarding instrument selection or development for young children should be made by individuals with the requisite programmatic and technical knowledge and after careful consideration of a variety of factors, including existing research, recommended practice, and available resources. Given the broad-based knowledge needed to make such decisions wisely, they cannot be made by a single individual or by fiat in legislation. Policy and legislation should allow for the adoption of new instruments as they are developed and validated.
- (I-10) Assessment tools should be constructed and selected for use in accordance with principles of universal design, so they will be accessible to, valid, and appropriate for the greatest possible number of children. Children with disabilities may still need accommodations, but this need should be minimized.
- (I-11) Extreme caution needs to be exercised in reaching conclusions about the status and progress of, as well as the effectiveness of programs serving, young children with special needs, children from language-minority homes, and other children from groups not well represented in norming or validation samples, until more information about assessment use is available and better measures are developed.

APPENDIX E: CCSSO Early Childhood Education SCASS Principles for Kindergarten Readiness Assessment

Council of Chief State School Officers' Early Childhood Education State Collaborative on Assessment and Student Standards, 2011, p. 2

Early child assessments conducted prior to, at the start of, and during kindergarten can be useful for a number of purposes if done well. Kindergarten readiness assessments should be used to directly support children's development and academic achievement to improve educational outcomes. To do so, kindergarten readiness assessment efforts should adhere to the following principles:

- Use multiple tools for multiple purposes.
- Address multiple developmental domains and diverse cultural contexts.
- Align with early learning guidelines and common core standards.
- Collect information from multiple sources.
- Implement in a systems-based approach.
- Avoid inappropriate use of assessment information, specifically including high-stakes decisions, labeling children, restricting kindergarten entry, and predicting children's future academic and life success.

nwea

Measuring What Matters™

NWEA is a not-for-profit organization that supports students and educators worldwide by providing assessment solutions, insightful reports, professional learning offerings, and research services. Visit NWEA.org to find out how NWEA can partner with you to help all kids learn.

January 2018 | KAP1607