

Preschool Policy Matters

Issue 7 / July 2004

NATIONAL
INSTITUTE
FOR
EARLY
EDUCATION
RESEARCH

Contact Us:

120 Albany Street, Suite 500
New Brunswick, NJ 08901

Tel (732) 932-4350

Fax (732) 932-4360

Website: nieer.org

E-mail: info@nieer.org

This policy brief is a joint publication of the National Institute for Early Education Research and the High/Scope Educational Research Foundation.



High/Scope Educational
Research Foundation
600 North River Street
Ypsilanti, MI 48198-2898, USA

Tel (734) 485-2000

Fax (734) 485-0704

Website: highscope.org

E-mail: info@highscope.org

THE STATE UNIVERSITY OF NEW JERSEY
RUTGERS

Preschool Assessment: A Guide to Developing a Balanced Approach

by Ann S. Epstein, Lawrence J. Schweinhart, Andrea DeBruin-Parecki and Kenneth B. Robin

Child assessment is a vital and growing component of high-quality early childhood programs. Not only is it an important tool in understanding and supporting young children's development, it is essential to document and evaluate program effectiveness. For assessment to be widely used though, it must employ methods that are feasible, sustainable and reasonable with regards to demands on budgets, educators and children.

Equally important, it must meet the challenging demands of validity (accuracy and effectiveness) for young children. It is the balance between efficiency and validity that demands the constant attention of policymakers — and an approach grounded in a sound understanding of appropriate methodology.

What We Know:

- Assessment is an ongoing process that includes collecting, synthesizing and interpreting information about pupils, the classroom and their instruction.
- Testing is one form of assessment that, appropriately applied, systematically measures skills such as literacy and numeracy.
- While it does not provide a complete picture, testing is an important tool, for both its efficiency and ability to measure prescribed bodies of knowledge.
- Alternative or "authentic" forms of assessment can be culturally sensitive and pose an alternative to testing, but they require a larger investment in establishing criteria for judging development and evaluator training.
- Child assessment has value that goes well beyond measuring progress in children – to evaluating programs, identifying staff development needs and planning future instruction.
- The younger the child, the more difficult it is to obtain valid assessments. Early development is rapid, episodic and highly influenced by experience. Performance on an assessment is affected by children's emotional states and the conditions of the assessment.

Policy Recommendations:

- Require that measures included in an assessment be selected by qualified professionals to ensure that they are reliable, valid and appropriate for the children being assessed.
- Develop systems of analyses so that test scores are interpreted as part of a broader assessment that may include observations, portfolios, or ratings from teachers and/or parents.
- Base policy decisions on an evaluation of data that reflects all aspects of children's development – cognitive, emotional, social, and physical.
- Involve teachers and parents in the assessment process so that children's behaviors and abilities can be understood in various contexts and cooperative relationships among families and school staff can be fostered.
- Provide training for early childhood teachers and administrators to understand and interpret standardized tests and other measures of learning and development. Emphasize precautions specific to the assessment of young children.

Purpose

This brief addresses the many questions about testing preschool children. Our purpose is three-fold: (a) to provide basic information about the terms and issues surrounding assessment; (b) to add an empirical and pragmatic perspective to what can sometimes be an impassioned debate; and (c) to support parents, policy makers and early childhood educators in using assessments to do what is best for young children and support the programs and policies that serve them.

Child assessment is a vital and necessary component of all high-quality early childhood programs. Assessment is important to understand and support young children's development. It is also essential to document and evaluate how effectively programs are meeting young children's educational needs, in the broadest sense of this term. For assessment to occur, it must be feasible. That is, it must meet reasonable criteria regarding its efficiency, cost, and so on. If assessment places an undue burden on programs or evaluators, it will not be undertaken at all and the lack of data will hurt all concerned. In addition to feasibility, however, assessment must also meet the demands of validity. The assessment must address the criteria outlined below for informing us about what children in real programs are learning and doing every day.

Efficiency and validity are not mutually exclusive but must sometimes be balanced against one another. The challenge is to find the best balance under the conditions that exist and when necessary, to work toward improving those conditions. Practically speaking, this means we must continue to serve children using research-based practices, fulfill mandates to secure program resources, and improve assessment procedures to better realize our ideal. This paper sets forth the criteria to be considered in striving to make early childhood assessment adhere to these highest standards.

Background

Concern with assessment in the early childhood field is not new. Decades of debate are summarized in the National Association for the Education of Young Children (NAEYC) publication *Reaching Potentials: Appropriate Curriculum and Assessment for Young Children*.¹ This position statement has been expanded by NAEYC and the National Association of Early Childhood Specialists in State Departments of Education (NAECS/SDE) in a new document titled *Early Childhood Curriculum, Assessment, and Program Evaluation: Building an Effective, Accountable System in Programs for Children Birth through Age 8*.²

What is new in this ongoing debate is the heightened attention to testing young children as a means of holding programs accountable for their learning. Peter Airasian's *Assessment in the Classroom* offers the following definitions:

"Assessment is the process of collecting, synthesizing, and interpreting information to aid classroom decision-making. It includes information gathered about pupils, instruction, and classroom climate."³

"Testing is a formal, systematic procedure for gathering a sample of pupils' behavior. The results of a test are used to make generalizations about how pupils would have performed on similar but untested behaviors."⁴

Testing is one form of assessment. It usually involves a series of direct requests to children to perform, within a set period of time, specific tasks designed and administered by adults, with predetermined correct answers. By contrast, alternative forms of assessment may be completed either by adults or children, are more open-ended, and often look at performance over an extended period of time. Examples include structured observations, portfolio analyses of individual and collaborative work, and teacher and parent ratings of children's behavior.

The current Head Start testing initiative focuses primarily on literacy and to a lesser extent numeracy. The rationale for this initiative, advanced in the No Child Left Behind Act and supported by the report of the National Reading Panel⁵, is that young children should acquire a prescribed body of knowledge and academic skills to be ready for school. Social domains of school readiness, while also touted as essential in a series of National Research Council reports⁶, are admittedly neither as widely mandated nor as "testable" as their academic counterparts. Hence, whether justified or not, they do not figure as prominently in the testing and accountability debate.

This brief responds to questions being asked of early childhood leaders about the use and misuse of testing for preschoolers 3 to 5 years old. This response is not merely a reactive gesture nor an attempt to advance and defend a specific position. Rather, the brief is intended as a resource to provide information about when and how preschool assessment in general, and testing and other forms of assessment in particular, can be appropriately used for purposes that include informing policy decisions about early childhood programming.

Continued next page

Continued from page 2

As a framework for providing this information, this policy brief accepts two realities. First, testing is, will be, and always has been, used to answer questions about the effectiveness of early childhood programs. Since early childhood programs attempt to increase children's knowledge and skills in specific content areas, evaluators have traditionally used testing, along with other assessment strategies, to determine whether these educational objectives have been achieved. Second, program accountability is essential, and testing is one efficient means of measuring it. Numerous research studies show that high-quality programs can enhance the academic and lifetime achievement of children at risk of school failure. This conclusion has resulted in an infusion of public and private dollars in early education. It is reasonable to ask whether this investment is achieving its goal. Testing can play a role in answering this accountability question.

With this reality as a background, we proceed to address two questions. First, given the current pervasive use of testing and its probable expansion, when and under what conditions can this type of assessment be used appropriately with preschool-age children? That is, what characteristics of tests and their administration will guarantee that we "do no harm" to children and that we "do help" adults acquire valid information? Second, given that even the most well-designed tests can provide only limited data, how can we maximize the use of non-test assessments so they add valuable information beyond that obtained through standardized testing procedures?

General Issues in Assessment

Uses of Child Assessment

Assessment can provide four types of information for and about children and their parents, teachers, and programs. Child assessment can:

1. Identify children who may be in need of specialized services. Screening children to determine whether they would benefit from specific interventions is appropriate when parents, teachers or other professionals suspect a problem. When screening indicates a problem, further assessments in several related domains are then usually administered to the child. In addition, data from parents and other adults involved with the child are considered in determining a diagnosis and course of treatment.

2. Plan instruction for individuals and groups of children.

Assessment data can be used by teachers to support the development of individual children, as well as to plan instructional activities for the class as a whole. In addition, information on developmental progress can and should be shared with parents to help them understand what and how their children are learning in the classroom and how they can extend this learning at home.

3. Identify program improvement and staff development needs.

Child assessments can provide formative evaluation data that benefit program and staff development. Findings can point to areas of the curriculum that need further articulation or resources or areas where staff need professional development. If children in the classroom as a whole are not making progress in certain developmental domains, it is possible that the curriculum needs revision or that teachers need some additional training. In conducting formative evaluations, child data are best combined with program data that measure overall quality, fidelity to curriculum implementation standards and specific teaching practices.

4. Evaluate how well a program is meeting goals for children.

It is this fourth purpose, sometimes called outcome or summative evaluation, that is the primary focus of this paper. Note that it is the program, not the young child, who should be held accountable through assessment. Although data may be collected on individual children, data should be aggregated to determine whether the program is achieving its desired outcomes. These outcomes may be defined by the program itself and/or by national, state, or district standards. How the outcomes are measured is determined by the inherent link between curriculum and assessment. Ideally, if a curriculum has clear learning objectives, those will drive the form and content of the measures. Conversely, thoughtful design of an appropriate assessment tool can encourage program developers to consider what and how adults should be teaching young children.



General Issues in Assessment *(continued)*

Reliability and Validity

Any formal assessment tool or method should meet established criteria for validity and reliability.⁷ *Reliability* refers to the consistency, or reproducibility of measurements. A sufficiently reliable test will yield similar results across time for a single child, even if different examiners or different forms of the test are used. Reliability is expressed as a coefficient between 0 (absence of reliability) and 1 (perfect reliability). Generally, for individualized tests of cognitive or special abilities, a reliability coefficient of .80 or higher is considered acceptable.

Validity is the degree to which a test measures what it is supposed to measure. Because tests are only valid for a specific purpose and assessments are conducted for so many different reasons, there is no single type of validity that is most appropriate across tests. *Content validity* refers to the extent to which the items on an instrument are representative of the key aspects of the domain it is supposed to measure. Irrelevant items or the absence of items to address some important element of a domain will negatively impact content validity. *Face validity* deals with appearance rather than content. A test has face validity if it appears to measure what it purports to measure.

In assessing young children, two aspects of validity have special importance – developmental validity and predictive validity. *Developmental validity* means the performance items being measured are developmentally suitable for the children being assessed. *Predictive validity* is the correlation between a test score and future performance on a relevant criterion. A test would have strong predictive validity, for example, if superior performance on the test was strongly associated with a high level of achievement later in school. The criterion to which test performance is compared may be another test or an indicator such as grade retention, special education placement or high school graduation. A test must be reliable in order to be valid but not all reliable tests are valid.

Principles and Recommendations for Early Childhood Assessments, a report to the National Education Goals Panel, noted that “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.”⁸ One prominent expert on early childhood assessment concludes, “research demonstrates that no more than 25 per cent of early academic or cognitive performance is predicted from information obtained from preschool or kindergarten tests.”⁹ Growth in the early years is rapid, episodic and highly influenced by environmental supports. Performance is influenced by children’s emotional and motivational states and by the assessment conditions themselves. Because these individual and situational factors affect reliability and validity, assessment of young children should be pursued with the necessary safeguards and caveats about the accuracy of the decisions that can be drawn from the results. These procedures and cautions are explored in the following.



Assessment Methods

The quality of an assessment depends in part upon decisions made before any measure is administered to a child. Before selecting an instrument for use with a given population of children, project designers should be able to explain why that specific measure is being used and what they hope to learn from the results. Selection of instruments is guided by the purposes and goals of the assessment. Assessment strategies lie along a continuum ranging from formal to informal. Types of measures that might be selected to represent either extreme include standardized testing (formal) and naturalistic observation (informal). The fundamental difference between formal and informal assessment is the degree of constraint placed on children's behavior, or level of intrusiveness into their lives.¹⁰

The ideal testing environment, as well as who is best qualified to administer measures, will depend in part on where along the formal-informal continuum an assessment lies. A standardized test is most effective when delivered by an examiner who has specialized training and experience with that specific instrument. Designers of standardized tests usually describe in test manuals the type of environment that must be created in order to obtain valid results. Most individual tests of cognitive ability must be administered in a controlled, relatively quiet area where a child is not likely to be distracted or interrupted. In contrast, informal assessments are ideally delivered by a child's teacher, or by another professional who interacts regularly with the child. These types of assessments often take place in a natural setting such as a classroom or playground. For the most part, examiners do not intrude in children's behavior when conducting an informal assessment.

The choice of an assessment strategy is also affected by the available resources in terms of time, money, and staff. Some assessments are more time and cost intensive than others. For example, one effective approach to identifying special needs (e.g., disabilities) is to use standardized tests to screen all children. These tools can be quickly and inexpensively administered to large populations of children. Children identified as potentially at risk or in need of further intervention can then receive follow-up evaluations using more intensive assessments including informal measures. Methods such as observation, parent interviews, analysis of work samples, or teacher ratings can lead to collection of in-depth and authentic data that reflect a "whole child" approach to the estimation of competence and need.

A comprehensive assessment normally requires a multi-method approach in order to encompass the many dimensions of children's skills and abilities. Formal and informal assessment strategies each have strengths and weaknesses, so an approach that combines or balances the two is most likely to provide a thorough evaluation of children across their cognitive, emotional, social, and biological strengths and needs. A repeated measures design is also preferable, especially with standardized tests, as performance of young children on assessment tasks will fluctuate according to mood and environment, as well as their rapid and sporadic development.

Standardized Testing

Standardized tests represent the most formal extreme of the assessment continuum because they place the greatest constraints on children's behavior. These tests are given under strictly controlled, standard conditions so that, to the extent possible, each child is assessed in exactly the same way. Standardized test scores allow for fair comparisons among individual or groups of test takers. Because standard administration is essential to obtain valid results, the skill of the examiner is of particular importance when using this type of assessment.

Standardized tests can be used to obtain information on whether a program is achieving its desired outcomes and are thus often integral components of systems of accountability. They are considered objective, time- and cost-efficient, and suitable for making quantitative comparisons of aggregated data across groups. Testing will only meet these expectations fully if the standard of comparison is developmentally and culturally appropriate. When used appropriately, standardized tests can effectively eliminate biases in assessment of individual children.

There is some concern about how well standardized tests work with young children. The younger the child, the more difficult it can be to obtain valid scores. Preschoolers may not understand the demands of the testing situation, and may respond unpredictably to the testing conditions. Performance is highly influenced by children's emotional states and experience, so that test scores across time may be relatively unstable. To address these limitations, examiners may choose to supplement standardized test scores with results from informal measures.

Assessment Methods *(continued)*

Informal Assessment Methods

Informal methods offer another approach to assessment. These other methods often fall under the banner of “authentic” or “naturalistic” assessments. They engage or evaluate children on tasks that are personally meaningful, take place in real life contexts, and are grounded in naturally occurring instructional activities. They offer multiple ways of evaluating students’ learning, as well as their motivation, achievement, and attitudes.

This type of assessment should be consistent with the goals, curriculum, and instructional practices of the classroom or program with which it is associated.¹¹ Authentic assessments do not rely on unrealistic or arbitrary time constraints, nor do they emphasize instant recall or depend on lucky guesses. Progress toward mastery is the key, and content is mastered as a means, not as an end.¹² To document accomplishments, assessments must be designed to be longitudinal, to sample the baseline, the increment, and the preserved levels of change that follow from instruction.¹³

Informal assessment can be more expensive than standardized testing. Like their counterparts in testing, informal measures must meet reasonable standards of demonstrated reliability and validity, though less emphasis tends to be placed on the psychometric quality of informal assessment tools. Their use, especially on a widespread scale, requires adequate resources. Assessors must be trained to acceptable levels of reliability. Data collection, coding, entry, and analysis are also time- and cost-intensive. This investment can be seen as reasonable and necessary, however, if the goal is to produce information about children’s competencies on real-life tasks in natural and authentic settings. Informal child assessment procedures that can meet acceptable levels of reliability and validity include observations, portfolios and ratings of children by teachers and parents.

Observations

In assessing young children, the principal alternative to testing is systematic observation of children’s activities in their day-to-day settings. Observation fits an interactive style of curriculum, in which give-and-take between teacher and child is the norm. Although careful observation requires effort, the approach has high ecological validity and intrudes minimally into what children are doing. Children’s activities naturally integrate all dimensions of their development—intellectual, motivational, social, physical, aesthetic, and so on.

Anecdotal notes alone, however, are not sufficient for good assessment. They do not offer criteria against which to judge the developmental value of children’s activities or provide evidence of reliability and validity. Instead, anecdotal notes should be used to complete developmental scales of proven reliability and validity. Such an approach permits children to engage in activities any time and anywhere that teachers can see them. It defines categories of acceptable answers rather than single right answers. It expects the teacher to set the framework for children to initiate their own activities. It embraces a broad definition of child development that includes not only language and mathematics but also initiative, social relations, physical skills and the arts. It is culturally sensitive when teachers are trained observers who focus on objective, culturally neutral descriptions of behavior (for example, “Pat hit Bob”) rather than subjective, culturally loaded interpretations (for example, “Pat was very angry with Bob”). Finally, it empowers teachers by recognizing their judgment as essential to accurate assessment.



Portfolios

One of the most fitting ways to undertake authentic, meaningful evaluation is through the use of a well-constructed portfolio system. Arter and Spandel define a portfolio as “a purposeful collection of student work that tells the story of the student’s efforts, progress, or achievement in (a) given area(s). This collection must include student participation in selection of portfolio content, the guidelines for selection, the criteria for judging merit, and evidence of student self-reflection.”¹⁴

Portfolios describe both a place (the physical space where they are stored) and a process. The process provides richer information than standardized tests, involves multiple sources and methods of data collection, and occurs over a representative period of time.¹⁵

In addition, they encourage two- and three-way collaboration between students, teachers, and parents; promote ownership and motivation; integrate assessment with instruction and learning; and establish a quantitative and qualitative record of progress over time.¹⁶ They can provide credible, meaningful evidence of students’ learning and development to parents, teachers, and others that can be used to inform practice and policy in the preschool classroom or at higher levels of the educational system.¹⁷

The purposes for which portfolios are used are as variable as the programs that use them.¹⁸ In some programs, they are simply a place to store the best work that has been graded in a traditional manner. In others, they are used to create longitudinal systems to demonstrate the process leading to the products and to design evaluative rubrics for program accountability. There are also programs that merely have students collect work that is important to them as a personal, non-evaluative record of their achievements. When portfolios are not used to judge ability in some agreed-upon fashion, they are usually not highly structured and may not even include reflective pieces that demonstrate student growth and understanding.

Portfolios are most commonly thought of as an assessment approach appropriate in elementary and secondary schools. Yet they have long been used in preschools to document and share children’s progress with parents, administrators and others. For portfolios to be used for program accountability, as well as student learning and reflection, the evaluated outcomes must be aligned with curriculum and instruction. Children must have some choice about what to include if they are to feel ownership and pride. Portfolios should document the

creative or problem-solving process as they display the product, encouraging children to reflect on their actions. Conversations with children about their portfolios engage them in the evaluation process and escalate their desire to demonstrate their increasing knowledge and skills. Sharing portfolios with parents can help teachers connect school activities to the home and involve parents in their children’s education.

Teacher Ratings

Teacher ratings are a way to organize teacher perceptions of children’s development into scales for which reliability and validity can be assessed. Children’s grades on report cards are the most common type of teacher rating system for older children. In the preschool years, teacher ratings are most commonly used to assess children’s social and emotional development. However, teacher ratings also can be used to assess children’s cognitive and language abilities. Teacher ratings can be specifically related to other types of child assessments including scores on standardized tests or other validated assessment tools, concrete and specific behavioral descriptions (e.g., frequency of participation in group activities, ability to recognize the letters in one’s name), or global assessments of children’s traits (e.g., cooperative, sociable, hard-working). Research shows that teacher ratings can have considerable short- and long-term predictive validity throughout later school years and even into adulthood.¹⁹

Parent Ratings

Parent ratings are a way to organize parent perceptions of children’s development into scales for which reliability and validity can be assessed. Soliciting parent ratings is an excellent way for teachers to involve them as partners in the assessment of their children’s performance. The very process of completing scales can inform parents about the kinds of behaviors and milestones that are important in young children’s development. It also encourages parents to observe and listen to their children as they gather the data needed to rate their performance. An example of the use of parent ratings is the Head Start Family and Child Experiences Survey (FACES) study, in which parents’ ratings of their children’s abilities and progress were related to measures of classroom quality and child outcomes.²⁰

Criteria of Reliable and Valid Preschool Assessment

Both the content and administration of measures must respect young children's developmental characteristics. Otherwise the resulting data will be neither reliable nor valid. Worse, the testing experience may be negative for the child. Further, the knowledge and skills measured in the testing situation must be transferable and applicable to real-world challenges that a child may face at home or at school. Otherwise the information gathered has no practical value. To produce meaningful data and minimize the risk of creating a harmful situation, all assessment tools for preschool-age children, whether formal or informal, should satisfy the following criteria:

- 1. Assessment should not make children feel anxious or scared.** It should not threaten their self-esteem or make them feel they have failed. Tests should acknowledge what children know—or have the potential to learn—rather than penalizing them for what they do not know. Examiners should be able to respond sensitively to each child's reactions to the testing situation.
- 2. Information should be obtained over time.** A single encounter, especially if brief, can produce inaccurate or distorted data. For example, a child may be ill, hungry, or distracted at the moment of testing. The test is then measuring the child's interest or willingness to respond rather than the child's knowledge or ability with respect to the question(s) being

asked. If time-distributed measurements are not feasible, then testers should note unusual circumstances in the situation (e.g., noise) or child (e.g., fatigue) that could render single-encounter results invalid and should either schedule a re-assessment or discount the results in such cases.

- 3. An attempt should be made to obtain information on the same content area from multiple and diverse sources, especially when repeated instances of data gathering are not feasible (e.g., due to time or budgetary constraints).** Just as young children have different styles of learning, so they will differentially demonstrate their knowledge and skills under varying modes of assessment. For example, a complete and accurate assessment of language ability may involve standardized tests, classroom observation and parent ratings. By measuring ability using multiple approaches, an assessment plan is also less likely to be individually or culturally biased.
- 4. The length of the assessment should be sensitive to young children's interests and attention spans.** The assessment period should probably not exceed 35–45 minutes. Further, testers should be sensitive to children's comfort and engagement levels, and take a break or continue the test at another time if the child cannot or does not want to proceed.
- 5. Testing for purposes of program accountability should employ appropriate sampling methods whenever feasible.** Testing a representative sample of the children who participate in a program avoids the need to test every child. Sampling strategies reduce the overall time spent in testing and minimize the chances for placing undue stress on children and burden on teachers and classrooms.



Other conditions that contribute to the reliability and validity of measures depend on the type of measure being used. Decisions on where testing should take place, who should administer the assessment, and the types of skills to be evaluated will differ for standardized tests and informal measures. For standardized test scores to be reliable and valid, the following criteria should be met:

1. Standardized tests should contain enough items to allow scores to represent a diverse range of individual ability.

In order to identify and distinguish among children of low, average and high levels of ability, standard scores must be applicable to children at either end of the spectrum and be sensitive to relatively minor differences in skill level.

2. Testing should take place in a controlled environment that at least approximates the conditions experienced by the population on which the measure was standardized.

Most tests need to be administered in a quiet area, relatively free from distraction. If testing is frequently interrupted or if a child's attention is drawn to other matters, results will not accurately reflect ability. Meeting environmental demands is particularly challenging with school-based assessments since space and privacy are at such a premium in schools.

3. Examiners should be appropriately trained and familiar with testing materials and procedures.

Because standard administration is the goal, examiners must understand the importance of considerations such as pacing, tone of voice, and establishing positive rapport with the child. Ideally, the examiner will be experienced and comfortable working with young children.

Creating a valid informal assessment for young children is a difficult task that demands unique considerations. It must be meaningful and authentic, evaluate a valid sample of behavior, be based on performance standards that are genuine benchmarks, and have authentic scoring. If scores on these measures are to resemble natural performance, it is incumbent upon the creators of informal assessment tools to design instruments that accomplish the following:

1. Informal assessments should take place in, or simulate, the natural environment in which the behavior being evaluated occurs.

It should avoid placing the child in an artificial situation. Otherwise, the assessment may measure the child's response to the setting rather than the child's ability to perform on the content.

2. The assessor should be knowledgeable regarding both the assessment materials and the children being assessed.

Ideally, the person administering the assessment is a teacher or another adult who interacts regularly with the child, so long as this familiarity does not invalidate the assessment through personal biases. When an outside researcher or evaluator must administer the assessment, it is best if the individual(s) spend time in the classroom beforehand, becoming a familiar and friendly figure to the children. Assessors who are not familiar with a child should learn what the child's typical interactions with adults are like.

3. Assessment should measure real knowledge in the context of real activities.

In other words, the assessment activities as well as the setting should not be contrived. They should resemble children's ordinary activities as closely as possible, for example, discussing a book as an adult reads it. Parent or teacher ratings should evaluate naturally occurring samples of behavior.

4. To the extent possible, assessments should be conducted as a natural part of daily activities rather than as a time-added or pullout activity.

Meeting this criterion helps to satisfy the earlier standards of a familiar place and assessor, especially if the assessment can be administered in the context of a normal part of the daily routine (for example, assessing book knowledge during a regular reading period). In addition, assessment that is integrated into standard routines avoids placing an additional burden on teachers or detracting from children's instructional time.



Conclusion

Recent years have seen a growing public interest in early childhood education. Along with that support has come the use of “high stakes” assessment to justify the expense and apportion the dollars. With so much at stake—the future of our nation’s children—it is imperative that we proceed correctly. Above all, we must guarantee that assessment reflects our highest educational goals for young children and neither restricts nor distorts the substance of their early learning. This brief sets forth the criteria for a comprehensive and balanced assessment system that meets the need for accountability while respecting the well-being and development of young children. Such a system can include testing, provided it measures applicable knowledge and skills in a safe and child-affirming situation. It can also include informal assessments, provided they too meet psychometric standards of reliability and validity. Developing and implementing a balanced approach to assessment is not an easy or inexpensive undertaking. But because we value our children and respect those charged with their education, it is an investment worth making.

Policy Recommendations

- Require that measures included in an assessment be selected by qualified professionals to ensure that they are reliable, valid and appropriate for the children being assessed.
- Develop systems of analyses so that test scores are interpreted as part of a broader assessment that may include observations, portfolios, or ratings from teachers and/or parents.
- Base policy decisions on an evaluation of data that reflects all aspects of children’s development—cognitive, emotional, social, and physical.
- Involve teachers and parents in the assessment process so that children’s behaviors and abilities can be understood in various contexts and so cooperative relationships among families and school staff can be fostered.
- Provide training for early childhood teachers and administrators to understand and interpret standardized tests and other measures of learning and development. Emphasize precautions specific to the assessment of young children.



Endnotes:

- ¹ Bredekamp, S., & Rosegrant, T. (Eds.) (1992). *Reaching Potentials: Appropriate Curriculum and Assessment for Young Children*. Washington, DC: National Association for the Education of Young Children.
- ² National Association for the Education of Young Children and 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: Authors. Available online at http://www.naeyc.org/resources/position_statements/pscape.asp.
- ³ Airasian, P. (2002). *Assessment in the classroom*. New York: McGraw-Hill.
- ⁴ Airasian, P. (2002).
- ⁵ National Reading Panel. (2000). *Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction*. Washington, DC: National Institute of Child Health and Human Development, National Institutes of Health.
- ⁶ National Research Council. (2000a). *Eager to learn: Educating our preschoolers*. Washington, DC: National Academy Press.; National Research Council. (2000b). *Neurons to neighborhoods: The science of early childhood development*. Washington, DC: National Academy Press.
- ⁷ American Educational Research Association, American Psychological Association, & National Council of Measurement in Education. (1999). *Standards for educational and psychological testing*. Washington, DC: American Psychological Association.
- ⁸ National Education Goals Panel. (1998). *Principles and recommendations for early childhood assessments*. Washington, DC: Author.
- ⁹ Meisels, S. (2003). Can Head Start pass the test? *Education Week*, 22(27), 44 & 29.
- ¹⁰ Hills, T. W. (1992). Reaching potentials through appropriate assessment. In S. Bredekamp & T. Rosegrant (Eds.), *Reaching potentials: Appropriate curriculum and assessment for young children* (Vol. 1, pp. 43-63). Washington, DC: National Association for the Education of Young Children.
- ¹¹ McLaughlin, M., & Vogt, M. (1997). *Portfolios in teacher education*. Newark, Delaware: International Reading Association.; Paris, S. G., & Ayers, L. R. (1994). *Becoming reflective students and teachers with portfolios and authentic assessment*. Washington DC: American Psychological Association.
- ¹² Wiggins, G. (1992). Creating tests worth taking. *Educational Leadership*, 49(8), 26-33.
- ¹³ Wolf, D., Bixby, J., Glenn, J., & Gardner, H. (1991). To use their minds well: Investigating new forms of student assessment. In G. Grant (Ed.), *Review of research in education*, Vol 17 (pp. 31-74). Washington D.C.: American Educational Research Association.
- ¹⁴ Arter, J. A., & Spandel, V. (1992). Using portfolios of student work in instruction and assessment. *Educational Measurement Issues and Practice*, 36-44.
- ¹⁵ Shaklee, B. D., Barbour, N. E., Ambrose, R., & Hansford, S. J. (1997). *Designing and using portfolios*. Boston: Allyn & Bacon.
- ¹⁶ Paris & Ayers (1994).; Paulson, F. L., Paulson, P. R., & Meyer, C. A. (1991). What makes a portfolio a portfolio? *Educational Leadership*, 48(5), 60-63.; Wolf, K., & Siu-Runyan, Y. (1996). Portfolio purposes and possibilities. *Journal of Adolescent and Adult Literacy*, 40(1), 30-37.; Valencia, S. W. (1990). A portfolio approach to classroom reading assessment: The whys, whats and hows. *The Reading Teacher*, 43(4), 338-340.
- ¹⁷ Herman, J. L., & Winters, L. (1994). Portfolio research: A slim collection. *Educational Leadership*, 52(2), 48-55.
- ¹⁸ Graves, D. H., & Sunstein, B. S. (1992). *Portfolio portraits*. New Hampshire: Heinemann.; Valencia, S.W. (1990).; Wolf, K. & Siu-Runyan, Y. (1996).
- ¹⁹ Schweinhart, L. J., Barnes, H. V., & Weikart, D. P. (1993). *Significant benefits: The High/Scope Perry Preschool study through age 27*. Ypsilanti, MI: High/Scope Press.
- ²⁰ Zill, N., Connell, D., McKey, R. H., O'Brien, R. et al. (2001). *Head Start FACES: Longitudinal Findings on Program Performance, Third Progress Report*. Washington, DC: Administration on Children, Youth and Families, U.S. Department of Health and Human Services.

**by Ann S. Epstein, Ph.D., Lawrence J. Schweinhart, Ph.D., Andrea DeBruin-Parecki Ph.D.,
and Kenneth B. Robin, PsyM.**

Ann S. Epstein, Ph.D., is Director of the High/Scope Early Childhood Division. Lawrence J. Schweinhart, Ph.D., is President of High/Scope. Andrea DeBruin-Parecki, Ph. D., is Director of the High/Scope Early Childhood Reading Institute. Kenneth B. Robin, Psy.M., is a Research Associate at the National Institute for Early Education Research.

Preschool Assessment: A Guide to Developing a Balanced Approach is issue 7 in a series of briefs, *Preschool Policy Matters*. This policy brief is a joint publication of the National Institute for Early Education Research and the High/Scope Educational Research Foundation. It may be reprinted with permission, provided there are no changes in the content.

Available online under “Resources/NIEER Publications” at nieer.org

This document was prepared with the support of The Pew Charitable Trusts. The Trusts’ *Starting Early, Starting Strong* initiative seeks to advance high quality prekindergarten for all the nation’s three-and four-year-olds through objective, policy-focused research, state public education campaigns and national outreach. The opinions expressed in this report are those of the authors and do not necessarily reflect the views of The Pew Charitable Trusts.



120 Albany Street, Suite 500 New Brunswick, New Jersey 08901
(Tel) 732-932-4350 (Fax) 732-932-4360

Website: nieer.org
Information: info@nieer.org



High/Scope Educational Research Foundation
600 North River Street Ypsilanti, MI 48198-2898, USA
(Tel) 734-485-2000 (Fax) 734-485-0704

Website: highscope.org
Information: info@highscope.org