

## **High/Scope Information Paper on Preschool Assessment**

by

Ann S. Epstein, Ph.D., Director, Early Childhood Division

Lawrence J. Schweinhart, Ph.D., President

Andrea DeBruin-Parecki, Ph. D., Director, Early Childhood Reading Institute

### **Purpose of Paper**

This paper addresses the many questions the High/Scope Educational Research Foundation has received about testing four-year-olds. Our reasons for sharing this paper with early childhood practitioners, policymakers, and parents 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 affirm our commitment to doing what is best for young children and supporting those who develop the programs and policies that serve them.

High/Scope believes 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 their 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 ecological 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* ecological validity are not mutually exclusive, but must sometimes be balanced against one another. Our challenge is to find the best balance under the conditions given and, when necessary, to work toward altering 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**

The 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* (Bredekamp & Rosegrant, 1992). This position statement has just been expanded 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* ([www.naeyc.org/resources/position\\_statements/pscape.asp](http://www.naeyc.org/resources/position_statements/pscape.asp)).

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. *Assessment in the Classroom* (Airasian, 2002) 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.

*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 in similar but untested behaviors.

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 objective observations, portfolio analyses of individual and collaborative work, and teacher and parent ratings of children's behavior.

The current 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 Panel (2000), 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 reports (notably *Eager to Learn*, 2000a and *Neurons to Neighborhoods*, 2000b), 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 information paper 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 paper 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 to inform policy decisions about early childhood programming.

As a framework for providing this information, High/Scope accepts two realities. First, testing is, will be, and in fact always has been, used to answer questions about the effectiveness of early childhood interventions. 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, this information paper proceeds to address two questions. First, given the current pervasive use of testing and its probably 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 too add valuable information over and above 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. In these cases, 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 child, who should be held accountable. 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 inextricable 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.

## **Reliability and Validity**

Any formal assessment tool or method should meet established criteria for validity and reliability (American Educational Research Association, American Psychological Association, & National Council of Measurement in Education, 1999). *Reliability* is defined as how well various measurements of something agree with each other, for example, whether a group of similar test items or two observers completing the same items have similar results. *Validity* has several dimensions. *Content* or *face validity* refers to how well an instrument measures what it claims to measure; *ecological validity* refers to the authenticity of the measurement context; and *construct validity* deals with the measure's conceptual integrity.

In assessing young children, two aspects of validity have special importance—developmental validity and predictive validity. *Developmental validity* means that the performance items being measured are developmentally suitable for the children being assessed. *Predictive validity* means the measure can predict children's later school success or failure, as defined by achievement test scores or academic placements (on-grade, retained in grade, or placed in special education) during the elementary grades. Over the longer term, predictive validity can even refer to such outcomes as adult literacy, employment, or avoiding criminal activity.

In *Principles and Recommendations for Early Childhood Assessments*, the National Education Goals Panel (1998) 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” (p. 5). Meisels (2003) claims “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” (p. 29). 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, the Panel recommended that assessment of young children 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 below.

## Testing

### Appropriate Uses of Testing

Standardized tests are used to obtain information on whether a program is achieving its desired outcomes. They are considered objective, time- and cost-efficient, and suitable for making quantitative comparisons. Testing can provide valid data when used appropriately and matched to developmental levels. Moreover, tests can act as teaching tools by providing a window into what children already know and where they need more time, practice, and/or help to improve.

Creating a valid assessment for young children is a difficult task. It must be meaningful and authentic, evaluate a valid sample of information learned, be based on performance standards that are genuine benchmarks, avoid arbitrary cut-off scores or norms, and have authentic scoring. The context for the test should be rich, realistic, and enticing (Wiggins, 1992). It is therefore incumbent upon the creators of assessment tools to design instruments that—unlike artificial drills—resemble natural performance. If these conditions are met, young children are more likely to recognize what is being asked of them, thus increasing the reliability and validity of the results.

### Criteria of Reliable and Valid Preschool Tests

Both the content and administration of tests 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 and perhaps the tester as well. Further, the knowledge and skills measured in the testing situation must be transferable and applicable in real-world settings. Otherwise the information gathered has no practical value. To produce meaningful data and minimize the risk of creating a harmful situation, tests for preschool-age children should satisfy the following criteria:

- 1. Tests should not make children feel anxious or scared.** They 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.
- 2. Testing should take place in, or simulate, the natural environment of the classroom.** It should avoid placing the child in an artificial situation. Otherwise, the test may measure the child's response to the test setting rather than the child's ability to perform on the test content.
- 3. Tests should measure real knowledge in the context of real activities.** In other words, the test activities as well as the test setting should not be contrived. They should resemble children's ordinary activities as closely as possible, for example, discussing a book as the adult reads it. Furthermore, tests should measure broad concepts rather than narrow skills, for example, alphabetic and letter knowledge sampled from this domain rather than familiarity with specific letters chosen by the adult.

- 4. *The tester should be someone familiar to the child.*** Ideally, the person administering the test would be a teacher or another adult who interacts regularly with the child. When an outside researcher or evaluator must administer the test, it is best if the individual(s) spend time in the classroom beforehand, becoming a familiar and friendly figure to the children. If this is not feasible, the appearance and demeanor of the tester(s) should be as similar as possible to adults with whom the child regularly comes in contact.
- 5. *To the extent possible, testing 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 tester, especially if the test 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, testing that is integrated into standard routines avoids placing an additional burden on teachers or detracting from children's instructional time.
- 6. *The 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.
- 7. *When repeated instances of data gathering are not feasible (e.g., due to time or budgetary constraints), an attempt should be made to obtain information on the same content area from multiple and diverse sources.*** 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 measure of letter knowledge may involve tests that employ both generative and recognition strategies.
- 8. *The length of the test should be sensitive to young children's interests and attention spans.*** If a test is conducted during a regular program activity (e.g., small-group time), the test should last no longer than is typical for that activity. If it is necessary to conduct testing outside regular activities, the assessment period should last 10–20 minutes. Further, testers should be sensitive to children's comfort and engagement levels, and take a break or continue the test at another day and time if the child cannot or does not want to proceed.
- 9. *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 and/or to administer all tests to any one child. Sampling strategies reduce the overall time spent in testing, and minimize the chances for placing undue stress on individual children or burdening individual teachers and classrooms.

## **Alternative Child Assessment Methods**

Alternative forms of assessment may be used by those who have reservations about, or want to supplement, standardized tests. These other methods often fall under the banner of “authentic” assessments. They engage children in 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 is consistent with the goals, curriculum, and instructional practices of the classroom or program with which it is associated (McLaughlin & Vogt, 1997; Paris & Ayres, 1994). 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 (Wiggins, 1989). 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 (Wolf, Bixby, Glenn & Gardener, 1991).

Alternative assessment can be more expensive than testing. Like their counterparts in testing, authentic measures must meet psychometric standards of demonstrated reliability and validity. 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 valid information. Alternative child assessment procedures that can meet the criteria of reliability and validity include observations, portfolios, and ratings of children by teachers and parents. These are described below.

### **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 (1991) 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 (p. 36).

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 (Shaklee, Barbour, Ambrose, & Hansford, 1997).

Portfolios have additional value. 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 (Paris & Ayres, 1994; Paulson, Paulson, & Meyer, 1991; Wolf & Siu-Runyan, 1996; Valencia, 1990). “Portfolios encourage teachers and students to focus on important student outcomes, provide parents and the community with credible evidence of student achievement, and inform policy and practice at every level of the educational system” (Herman & Winters, 1994, p. 48).

The purposes for which portfolios are used are as variable as the programs that use them (Graves & Sunstein, 1993; Valencia, 1990; Wolf & Siu-Runyan, 1996). In some programs, they are simply a place to store 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 alternative assessments 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 in order 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 engages them in the evaluation process and escalates 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. When completed objectively, report-card grades are tied to students' performance on indicators with delineated scoring criteria, such as examinations or projects evaluated according to explicitly defined criteria. In these ways, 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 (Schweinhart, Barnes, & Weikart, 1993).

### **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 (Zill, Connell, McKey, O'Brien et al., 2001).

### **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 paper sets forth the criteria for a comprehensive and balanced assessment system that meets the need for accountability while respecting the welfare 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 alternative 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 care, it is an investment worth making.

## References

- Airasian, P. (2002). *Assessment in the classroom*. New York: McGraw-Hill.
- 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.
- Arter, J. A., & Spandel, V. (1992). Using portfolios of student work in instruction and assessment. *Educational Measurement Issues and Practice*, 36–44.
- Bredenkamp, S., & Rosegrant, T. (Eds.) (1992). *Reaching Potentials: Appropriate Curriculum and Assessment for Young Children*. Washington, DC: National Association for the Education of Young Children.
- Graves, D. H., & Sunstein, B. S. (1992). *Portfolio portraits*. New Hampshire: Heinemann.
- Herman, J. L., & Winters, L. (1994). Portfolio research: A slim collection. *Educational Leadership*, 52(2), 48–55.
- McLaughlin, M., & Vogt, M. (1997). *Portfolios in teacher education*. Newark, Delaware: International Reading Association.
- Meisels, S. (2003, 19 March). Can Head Start pass the test? *Education Week*, 22(27), 44 & 29.
- National Association for the Education of Young Children and National Association of Early Childhood Specialists in State Departments of Education (2003, November). *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 [www.naeyc.org/resources/position\\_statements/pscape.asp](http://www.naeyc.org/resources/position_statements/pscape.asp).
- National Education Goals Panel. (1998). *Principles and recommendations for early childhood assessments*. Washington, DC: Author.
- 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.
- 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.
- Paris, S. G., & Ayers, L. R. (1994). *Becoming reflective students and teachers with portfolios and authentic assessment*. Washington DC: American Psychological Association.
- Paulson, F. L., Paulson, P. R., & Meyer, C. A. (1991). What makes a portfolio a portfolio? *Educational Leadership*, 48(5), 60–63.
- 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.
- Shaklee, B. D., Barbour, N. E., Ambrose, R., & Hansford, S. J. (1997). *Designing and using portfolios*. Boston: Allyn & Bacon.
- Valencia, S. W. (1990). A portfolio approach to classroom reading assessment: The whys, whats and hows. *The Reading Teacher*, 43(4), 338–340.
- Wiggins, G. (1992). Creating tests worth taking. *Educational Leadership*, 49(8), 26–33.
- 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.
- Wolf, K., & Siu-Runyan, Y. (1996). Portfolio purposes and possibilities. *Journal of Adolescent and Adult Literacy*, 40(1), 30–37.
- Zill, N., Connell, D., McKey, R. H., O'Brien, R. et al. (2001, January). *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.

## High/Scope Assessment Resources

High/Scope has developed and validated three preschool assessment instruments. Two are for children, one focusing specifically on literacy and the other more broadly on multiple domains of development. The third measure is used to assess and improve the quality of all aspects of early childhood programs. These alternative assessments are described below.

### Early Literacy Assessment

In the Fall of 2004, High/Scope will release the *Early Literacy Assessment* (ELA), which will evaluate the four key principles of early literacy documented in the Early Reading First Grants and the No Child Left Behind legislation: phonological awareness, alphabetic principle, comprehension, and concepts about print. Evaluation will take place in a meaningful context that is familiar to children – book reading with an adult – and will include measures of upper- and lowercase alphabet letters, rhyming pairs, alliterative word pairs, book orientation and directionality, and opportunities to predict or recall information. The ELA can be used by programs implementing the new *High/Scope Growing Readers Curriculum* and *Reading Families*, developed by the High/Scope Early Childhood Reading Institute. This literacy assessment is also appropriate for use by all early childhood settings implementing and evaluating other developmentally based early reading curricula.

High/Scope Educational Research Foundation. (In press). *Early Literacy Assessment*. Ypsilanti, MI: High/Scope Press.

High/Scope Educational Research Foundation. (In press). *Growing Readers*. Ypsilanti, MI: High/Scope Press.

High/Scope Educational Research Foundation. (In press). *Reading Families*. Ypsilanti, MI: High/Scope Press.

### Child Observation Record

High/Scope has just published the second edition of the *Preschool Child Observation Record* (COR; High/Scope, 2003). The new version has been updated to take into account the current assessment needs of early childhood educators (for example, sections on language and literacy and math and science foundations have been expanded) and to reflect the many ideas gained from over a decade of experience with the earlier version (High/Scope, 1992). High/Scope first developed and validated the COR for use in all developmentally appropriate early childhood programs, regardless of the curriculum model used, with funding from the federal Head Start Bureau. At that time, a two-year study with Head Start teaching teams found the COR to be a feasible, valid, and reliable assessment tool (Schweinhart, McNair, Barnes, & Lerner, 1993). To develop the second edition, Head Start teachers of the Equal Opportunity Commission of St. Clair County Children's Services in Port Huron, Michigan, collected data that allowed High/Scope to confirm the reliability and validity of the revised assessment tool.

To use the COR, a trained teacher assesses each child's behavior and activities on 32 items in six categories of development:

- *Initiative*—making choices and plans, solving problems with materials, initiating play, taking care of personal needs
- *Social relations*—relating to adults, relating to other children, resolving interpersonal conflict, and understanding and expressing feelings
- *Creative representation*—making and building models, drawing and painting pictures, pretending
- *Movement and music*—moving in various ways, moving with objects, feeling and expressing steady beat, moving to music, and singing
- *Language and literacy*—listening to and understanding speech, using vocabulary, using complex patterns of speech, showing awareness of word sounds, demonstrating knowledge about books, using letter names and sounds, reading, and writing
- *Mathematics and science*—sorting by attributes, identifying patterns, comparing properties, identifying position and direction, identifying sequence, change and causality, identifying materials and properties, and identifying natural and living things

Over several months, teachers write brief notes describing episodes of young children’s behavior in these domains, occurring during the course of normal program activities. Teacher then use these notes to rate individual children’s behavior on the 32 five-level COR items. For example, the item “making choices and plans” has the following five assessment levels:

- (1) Child indicates a choice by pointing or some other action.
- (2) Child expresses a choice in one or two words.
- (3) Child expresses a choice with a short sentence.
- (4) Child makes a plan with one or two details.
- (5) Child makes a plan with three or more details.

Teachers score the COR two or three times a year, obtaining initial ratings six to eight weeks into the program year. The COR has also been used by trained outside observers, observing sets of four children over three program sessions, for research on programs themselves (Epstein, 1993).

The COR and similar development efforts point the way to a style of early childhood assessment that respects both the methods of the assessment field and the process and outcome goals of the early childhood field. Without such assessment tools, the early childhood field runs the grave risk of distortion by tests that embody a style of education that is inconsistent with developmentally appropriate practice. With such assessment tools, early childhood programs can realize the great promise that society hopes for and expects them to accomplish.

- Epstein, A. S. (1993). *Training for quality: Improving early childhood programs through systematic inservice training*. Ypsilanti, MI: High/Scope Press.
- High/Scope Educational Research Foundation. (1992). *The High/Scope Child Observation Record for Ages 2½ to 6*. Ypsilanti, MI: High/Scope Press.
- High/Scope Educational Research Foundation. (2003). *The Preschool Child Observation Record, Second Edition*. Ypsilanti, MI: High/Scope Press.
- Schweinhart, L. J., McNair, S., Barnes, H., & Larner, M. (1993, Summer). Observing young children in action to assess their development: The High/Scope Child Observation Record study. *Educational and Psychological Measurement*, 53, 445–455.

## Program Quality Assessment

The combination of child and program assessment provides valuable data for staff development and program enhancement. High/Scope recently published the second edition of the *Preschool Program Quality Assessment* (PQA; High/Scope, 2003), a validated and comprehensive measure of program quality that can be used for these purposes. The PQA reflects research-based knowledge about “best practices” in the field as a whole, and is therefore appropriate for use in all center-based settings, not just those using the High/Scope approach. The new PQA updates the earlier version (High/Scope, 1998) and is particularly cognizant of the educational strategies that support early learning. It is aligned with the Head Start Performance Standards and can also be aligned with state program standards.

The PQA can be administered as a self-assessment instrument or completed by outside program evaluators. Using a combination of observation and standardized interview techniques, trained PQA raters compile objective anecdotal evidence and score programs on a series of 63 five-point scales. The endpoints and the midpoint of each scale are defined by concrete indicators and examples to assure reliable and valid ratings. The items are grouped under the seven dimensions listed below to provide a comprehensive profile of program quality. The first four sections are completed at the classroom level; the last three sections are completed at the agency level.

- *Learning Environment* looks at the setting’s compliance with health and safety standards, the arrangement and equipping of the indoor and outdoor program space, and the variety, type, and diversity of learning materials.
- *Daily Routine* looks at the consistency of the program’s schedule and the diversity of individual and group experiences it provides for children, especially those shown by research to enhance their intellectual, social-emotional, and physical development.
- *Adult-Child Interaction* looks at how adults establish a warm and secure atmosphere that supports young children’s communication and language development, play initiatives, peer interactions, problem-solving, and overall academic and social learning.
- *Curriculum Planning and Assessment* looks at the program’s implementation of a consistent curriculum model, its procedures to plan for individual children as well as group activities, its record-keeping system, and its use of validated child assessment measures.
- *Parent Involvement and Family Services* looks at the nature and diversity of opportunities for parent involvement, staff-family communication, and provision of diagnostic and special education services, referral to community services, and transition to kindergarten.
- *Staff Qualifications and Staff Development* looks at the education and experience of staff (program director, teachers, support staff) and provision of ongoing inservice training using strategies shown by research to be effective mechanisms for professional development.
- *Program Management* looks at compliance with local licensing requirements, continuity in instructional staff, procedures for ongoing program assessment and improvement, recruitment and enrollment procedures, operational policies, accessibility to those with disabilities, and adequacy of funding to carry out all aspects of program operations.

In a series of validation studies (e.g., Epstein, 1993 & 1999; Jurkiewicz, 2003), scores on the PQA have been significantly associated with staff qualifications and training on the one hand, and young children's developmental progress on the other. By relating PQA scores to children's performance on assessment measures, administrators and evaluators can determine which programmatic areas are critical to achieving curriculum objectives. Staff development initiatives can then be targeted to those program components that will be most effective in enhancing children's development.

Epstein, A. S. (1993). *Training for quality: Improving early childhood programs through systematic inservice training*. Ypsilanti, MI: High/Scope Press.

Epstein, A. S. (1999). Pathways to quality in Head Start, public school, and private non-profit early childhood programs. *Journal of Research in Childhood Education*, 13(2), 101–119.

High/Scope Educational Research Foundation. (1998). *The High/Scope Program Quality Assessment: Preschool Version*. Ypsilanti, MI: High/Scope Press.

High/Scope Educational Research Foundation. (2003). *The Preschool Program Quality Assessment, Second Edition*. Ypsilanti, MI: High/Scope Press.

Jurkiewicz, T. (2003). *The Revised Preschool PQA: Report on psychometric properties*. Instrument evaluation report to the Michigan Department of Education. Ypsilanti, MI: High/Scope Educational Research Foundation, Research Division.

To find out more about or to order High/Scope assessment instruments, visit our online catalog at [www.highscope.org](http://www.highscope.org) or contact the High/Scope Press toll-free by telephone (1-800-407-7377) or fax (1-800-442-4329).