Preschool Assessment: What’s Right for Four-Year-Olds

by Ann Epstein, Ph.D., Director, Early Childhood

In recent years there has been a growing public investment in early childhood education. Along with taxpayer support has come enhanced scrutiny to justify the expenditures on these programs. News items about testing four-year-olds in Head Start or state-funded pre-kindergarten programs appear in features covering the federal No Child Left Behind (NCLB) act. While concerns about assessing young children are not new, high-stakes testing to hold early childhood programs accountable is a recent phenomenon. In response to this heightened public and professional attention, The High/Scope Information Paper on Preschool Assessment, on which this article is based, provides practitioners, policymakers, and parents with evidence-based guidance on how to obtain meaningful data on young children and the programs that serve them.

High/Scope believes that assessment is a vital and necessary component of all high-quality early childhood programs. Assessment is important to understanding and supporting young children’s development; it is also essential to documenting and evaluating how effectively programs are meeting children’s broadest educational needs—cognitive, social, emotional, and physical. Assessment must be feasible; that is, it must be reasonable to administer in terms of efficiency, cost, and so on. It must also be valid; that is, it must measure what children are really learning and doing every day in their program environments. Feasibility and validity are not mutually exclusive, but sometimes they must be balanced against each other. The High/Scope paper sets forth the criteria for meeting the highest research standards in a manner that is sensitive to the well being of the children who undergo assessments and the adults who administer them.

Definitions

Assessment is the process of collecting, synthesizing, and interpreting information about children and classrooms to aid decision making. Testing is one form of assessment. It usually involves a series of direct requests to children to perform specific tasks, designed and administered by adults, given within a set period of time, and having predetermined correct answers. By contrast, alternative forms of assessment may be completed by either 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.

Any assessment method must be reliable and valid. Reliability is defined as how well various measures agree with one other, for example, whether similar test items or trained independent observers produce comparable results. In assessing young children, two aspects of validity have special importance. Developmental validity means the items are developmentally suitable for that age group. Predictive validity means the measure can predict children’s school success (achievement, grade placement, graduation), and such later outcomes as adult literacy, employment, or avoidance of criminal activity.

Uses of Assessment

Assessment provides four types of information for and about children and their parents, teachers, and programs. It can be used to

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, data from specialized assessments plus information from the adults involved with the child can help to determine a diagnosis and plan of action to address the difficulties.

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 how to extend classroom learning in the home.

3. Identify program improvement and staff development needs. Child assessments can provide evaluation data for improving programs and teach-
ing practices. They can point to curriculum areas that need to be expanded or that should be the focus of professional development for staff.

4. Evaluate how well a program is meeting goals for children. This fourth purpose, sometimes called outcome or summative evaluation, is the primary focus of the High/Scope paper on preschool assessment. The paper emphasizes that it is the program, not the child, that should be held accountable. Although data may be collected on individual children, they should be aggregated to determine whether the program is achieving its desired outcomes according to national, state, or local standards.

Criteria of Reliable & Valid Preschool Tests

When used appropriately, standardized tests can provide valid data on whether a program is achieving its desired outcomes. Moreover, these 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.

When it is necessary to administer standardized tests to young children, both the content and procedures for administering the tests should meet 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. Tests should not be administered in an artificial situation. If they are, they 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. The test activities as well as the test setting should resemble children's ordinary activities as closely as possible. Furthermore, tests should measure broad concepts rather than narrow skills. For example, alphabetic and letter knowledge should be sampled from all age-appropriate possibilities in this area rather than from 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 outsider must administer the test, it is best if that person spends 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 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 satisfy the previous 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. Testing that is integrated into everyday routines also avoids placing an additional burden on teachers or detracting from children's instructional time.

6. The information should be obtained over time. A single brief testing session can produce inaccurate or distorted data, especially if a child is ill, hungry, or distracted during testing. If time-distributed measurements are not feasible, testers should note unusual circumstances in the situation (such as noise) or child (such as fatigue) that could render the results invalid, and either schedule a reassessment or discount the findings entirely.

7. When repeated instances of data gathering are not feasible (for example, 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 demonstrate their knowledge and skills differently when assessed through varying methods. For an accurate picture of a child's true abilities, multiple forms of assessment are essential.

8. The length of the test should accommodate young children's interests and attention spans. If a test is conducted during a regular program activity (such as 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. Testers should be sensitive to children's comfort and engagement levels, taking breaks or distributing the test over more than one day if needed.

9. Testing for purposes of program accountability should employ appropriate sampling methods whenever feasible. Testing a representative sample of children avoids the need to test every child in a program or to administer all the tests to any one child. Sampling strategies reduce the overall time spent in testing and minimize the
stress and burden placed on individual children and teachers.

**Alternative Child Assessment Methods**

Alternative forms of assessment—observations, portfolios, teacher and parent ratings—may be used instead of, or in addition to, standardized tests. Widespread application of these methods requires adequate resources to train assessors to acceptable levels of reliability and validity, to collect information, and to code and analyze the data. They can be expensive. Yet, they are worth the investment if the goal is to produce valid information for making programmatic and funding decisions.

**High/Scope Assessment Resources**

High/Scope has developed two validated assessment instruments for early childhood programs and is field-testing a third. The second edition of the *Preschool Child Observation Record* (COR; High/Scope, 2003) has just been published. This updated COR has expanded sections on language and literacy as well as math and science and reflects the many insights gained from over a decade of experience with the earlier version. The combination of child and program assessment provides valuable data for staff development and program enhancement. High/Scope also recently published the second edition of the *Preschool Program Quality Assessment* (PQA; High/Scope, 2003), which reflects best practices in the field as a whole and can therefore be used in any center-based program, not just those using the High/Scope approach. The new PQA is aligned with the Head Start Performance Standards and can also be aligned with state program standards. Finally, High/Scope’s Early Childhood Reading Institute is currently field-testing the *Early Literacy Assessment* (ELA) for a Fall 2004 release. The ELA uses a familiar context—book reading with an adult—to assess children on the four key principles required by Early Reading First and NCLB: phonological awareness, alphabetic principles, comprehension, and print concepts.

**Conclusion**

The pressures inherent in high-stakes testing demand that we adopt a reasonable approach to holding early childhood programs accountable for outcomes. The *High/Scope Information Paper on Preschool Assessment* sets forth the criteria for a comprehensive system—including tests as well as alternative assessment procedures—that respects the welfare and development of young children. Implementing a balanced approach to assessment is not an easy or inexpensive undertaking. But because we value our children and those charged with their care, it is an investment worth making.

References


To download a complete copy of The High/Scope Information Paper on Preschool Assessment, visit the High/Scope Web site at www.highscope.org and click “Testing 4-Year-Olds” under High/Scope Headlines, or go directly to http://www.highscope.org/Assessment/assess-stmt.pdf. To learn more about High/Scope assessment instruments, visit the online catalog (www.highscope.org/welcome.asp) or contact the High/Scope Press order line toll-free at 1-800-40-PRESS (telephone) or 1-800-442-4329 (fax).