



Eager to Learn: Educating Our Preschoolers

Barbara T. Bowman, M. Suzanne Donovan, and M. Susan Burns, Editors; Committee on Early Childhood Pedagogy; National Research Council

ISBN: 0-309-50389-2, 468 pages, 6 x 9, (2000)

This PDF is available from the National Academies Press at:
<http://www.nap.edu/catalog/9745.html>

Visit the [National Academies Press](http://www.nap.edu) online, the authoritative source for all books from the [National Academy of Sciences](http://www.nap.edu), the [National Academy of Engineering](http://www.nap.edu), the [Institute of Medicine](http://www.nap.edu), and the [National Research Council](http://www.nap.edu):

- Download hundreds of free books in PDF
- Read thousands of books online for free
- Explore our innovative research tools – try the “[Research Dashboard](#)” now!
- [Sign up](#) to be notified when new books are published
- Purchase printed books and selected PDF files

Thank you for downloading this PDF. If you have comments, questions or just want more information about the books published by the National Academies Press, you may contact our customer service department toll-free at 888-624-8373, [visit us online](#), or send an email to feedback@nap.edu.

This book plus thousands more are available at <http://www.nap.edu>.

Copyright © National Academy of Sciences. All rights reserved.

Unless otherwise indicated, all materials in this PDF File are copyrighted by the National Academy of Sciences. Distribution, posting, or copying is strictly prohibited without written permission of the National Academies Press. [Request reprint permission for this book](#).



Assessment in Early Childhood Education

THE USE OF TESTS AND ASSESSMENTS¹ as instruments of education policy and practice is growing. Throughout the school years, tests are used to make decisions about tracking, promotion or retention, placement, and graduation. Many teachers use tests or assessments to identify learning differences among students or to inform instructional planning. Widespread public concern to raise education standards has led states increasingly to use large-scale achievement tests as instruments of accountability (National Research Council, 1999a). Given their prevalence in the education system as a whole, it is not entirely surprising that the use of tests

¹Although the terms are not mutually exclusive, the word “test” tends to be used to refer to standardized instruments, formally administered, and designed to minimize all differences in the conditions of testing presented to test takers. There are both individually administered and group-administered standardized tests. The group-administered multiple-choice format is what people often have in mind when the term is used. Assessments embrace a wide array of formats (observations, performance measures, portfolios, essays). The term “assessment” is often used to communicate the intention to build a richer picture of the ways in which people think, learn, and work. They frequently are conducted over a longer period of time than group tests permit. Standardized tests focus on individual differences, answering the question “How does this individual compare with all others in the reference population?” Assessments reflect the interest of modern cognitive theory in the processes of learning and knowing in a given individual.

and assessments is increasingly common in preschool settings as well.

In the current early childhood education milieu, there are four primary reasons for assessment (Shepard et al., 1998):

- Assessment to support learning,
- Assessment for identification of special needs,
- Assessment for program evaluation and monitoring trends, and
- Assessment for school accountability.

Assessment to support learning, the first and most important of these purposes, refers to the use of assessments to provide teachers with information that can serve as a basis for pedagogical and curriculum decisions. Information presented in earlier chapters—about early learning, about the episodic course of development in any given child and the enormous variability among young children in background and preparation for school, about the centrality of adult responsiveness to healthy cognitive and emotional development—leads to the conclusion that what preschool teachers do to promote learning needs to be based on *what each child brings to the interaction*. Assessment broadly conceived is a set of tools for finding this out. The second reason for assessing young children is to diagnose suspected mental, physical, or emotional difficulties that may require special services. The final two purposes can be combined under the rubric of assessment to make policy decisions.

Each of these purposes represents an important opportunity for test or assessment data to inform judgment—if the tests or assessments are used carefully and well. No single type of assessment can serve all of these purposes; the intended purpose will determine what sort of assessment is most appropriate. There is much to be learned from the experience in other educational settings about the uses, misuses, and unintended consequences of testing (e.g., Haertel, 1989; Gifford, 1993; National Research Council, 1982a; U.S. Congress Office of Technology Assessment, 1992; Shepard, 1991). And there is much to remember about the developmental status of young children, including the nascent state of their attention and self-regulation abilities, that makes as-

assessment even more challenging than in other populations. The psychometric models on which testing has traditionally been based make standardized tests particularly vulnerable to misinterpretation (Shepard et al., 1998).

Experts agree on a number of guiding principles that apply to any setting in which tests or assessments are used in decision processes (see, for example, American Education Research Association, American Psychological Association, National Council on Measurement in Education, *Standards for Educational and Psychological Testing*, 1999; Shepard et al., 1998; National Research Council, 1999a). Perhaps the most important of these is that no single procedure should be the sole basis for decisions, or, put positively, important educational decisions should be grounded in multiple sources of information. These might include individual assessments of various sorts, standardized tests, observation, investigation of social and cultural background, and interviews. A corollary of this statement is that no test score should be looked on as infallible or immutable.

A second point of consensus is the requirement of measurement validity. Whether test or assessment, formal or informal, criterion- or norm-referenced, the measures being used need to have a reasonable level of accuracy. This means that school officials and teachers must inform themselves. They need to understand the strengths and weaknesses of various assessment approaches for the purposes they have in mind. They need to know what the research says about the specific instruments they intend to use. They need to develop sophistication in the interpretation of the information gleaned from tests and assessments.

A third important principle is borrowed from the Hippocratic oath to first do no harm. When test or assessment information is used for placement, school readiness, or other high-stakes decisions, it behooves educators to pay attention to the consequences and to make sure that they are educationally beneficial.

ISSUES IN STANDARDIZED ASSESSMENT OF YOUNG CHILDREN

Beyond these principles that apply generally to educational testing and assessment, there are important considerations that

become particularly salient in the assessment of young children. The evolution of views on the optimum conditions for assessment provides a good example. The traditional psychometric concerns with standardization have in the past been applied to assessments of young children. Individual or group tests were administered under controlled circumstances in highly structured environments that were as similar to one another as possible. But dissatisfaction among many early childhood professionals concerning the conventional model of norm-referenced assessment has in recent years brought a shift in emphasis toward conducting assessments in settings that are comfortable, familiar, nonthreatening, and of interest to the child (see Meisels and Provence, 1989; Greenspan and Wieder, 1998). There is evidence that such settings better enable young children to show what they know, what they can do, and what they are experiencing (Meisels, 1996b).

Many of the reasons that can be advanced to support this approach to assessment environments (among them the motivation to design assessments that have greater ecological validity) could also pertain to assessment of older children and adults. But there are also developmental and cultural characteristics of young children that can be attended to more effectively in more flexible settings than is possible in most standardized testing environments (Bracken, 1987). Examiner-examinee rapport, for example, is much trickier with very young children simply because of their very limited experience; race, gender, culture—even size—can significantly influence the child's ability to focus and attend. The motivation, state of arousal, and disposition of the very young child are likely to be much more variable than is the case for older test takers, who have more developed self-regulation abilities. The very young are by definition less familiar with the whole notion of and materials used for assessment, so that creating a more flexible and responsive environment that promotes the physical and emotional comfort of the child is likely to produce a more accurate picture of the child's knowledge, skills, achievement, or personality (Meisels, 1994).

Developmental Considerations

Young children have, in varying degrees, developmental limitations on several important (and often unrecognized) dimensions

that can affect assessment. We have made reference above to the nascent state of the ability to focus and attend in children of the ages of concern in this report. Likewise, the capacity to be purposeful and intentional, although undergoing rapid development, is certainly less than fully formed. In assessment situations, therefore, young children often have difficulty attending to verbal instructions, situational clues, or other instructions and stimuli. They may have difficulty understanding the demand characteristics of the measurement situation, and they may not be able to control their behavior sufficiently to meet these demands (Gelman and Gallistel, 1978).

Obviously, there are also implications for assessment of the emergent state of young children's verbal abilities. Depending on the child's functional capacity to use ideas and to communicate thoughts and feelings, for example, examiners may need to make inferences based on the child's overt motor behaviors or parent report, rather than direct response. Observational modes of assessment and interviews lend themselves to this situation. And although tests or assessments that require examiners to elicit responses can be useful, for example, to assess the child's grasp of key concepts about written language, math, and science, a different view is provided when the assessment casts the child as the initiator. Elicited language in particular may be qualitatively different from language that is used functionally in everyday contexts, and thus not representative of the child's functioning.

Cultural Considerations

In an important sense, education can be viewed as the journey from natal culture to school culture to the culture of the larger society. Education inevitably involves cognitive socialization, that is, learning the repertoires of cognitive skills that are required for successful functioning in the dominant culture. A modern industrial society like the United States that is technologically advanced, as Ogbu (1994) puts it, will possess a repertoire of cognitive skills appropriate for advanced technological culture. Technological intelligence is appropriate to and a prerequisite for functioning competently in that culture.

In a highly heterogeneous society such as ours, child care centers and preschools are in a position to play an extremely impor-

tant role in helping youngsters get off to a good start on that journey. But that requires teachers to be sensitive to the influences of culture both in choosing pedagogical strategies and in the use and interpretation of assessments. There are any number of obvious pitfalls that teachers are well aware of, for example, the use of English-language assessments that depend on verbal interactions with children who are growing up surrounded by a different home language. But valid assessment requires being aware of much more subtle factors as well. For example, there are great cultural variations in the ways in which adults and children communicate (National Research Council, 1999b:96-101). Ethnographic research has shown striking differences in how adults and children interact verbally. Many American Indian and African American subcultures do not cultivate the role of information giver that characterizes American middle-class children; the young are expected to learn through quietly observing adults (Heath, 1983). In some communities, children are seldom direct conversational partners with adults; children eavesdrop on adults, while older children take on the task of directly teaching social and intellectual skills (Ward, 1971). Children from these cultural backgrounds are not nearly as likely to show their actual verbal ability in assessment situations based on the elicited response model as those for whom question and answer is a familiar ritual. Culture also plays a role in determining which cues are most salient to children (Rogoff, 1990).

One of the greatest dangers in assessing young children is to associate developmental status with the norms of the dominant middle-class culture. This will lead to misunderstanding of children's functional abilities and misjudging pedagogical strategies. To draw again on Heath's ethnographic studies (1981, 1983), white, middle-class mothers begin questioning games from earliest infancy—"Where is Teddy Bear? Ah, there he is." Children exposed to these "known-answer" rituals are more likely than others to be comfortable with the question-and-answer dialogues typically encountered in preschool and school settings. Chapter 5 emphasized the importance of having a toolkit of teaching strategies, with each tool serving different ends and none being most effective for all purposes. The same can be said of assessment.

Sensitivity to the child's current competence means taking the child's home culture into account in assessment.

Assessing Children with Disabilities

One of the most difficult issues in early childhood assessment has to do with children who appear to need special assistance as a result of cognitive, emotional, visual, auditory, or motor impairments. On one hand, research has demonstrated that early intervention can often reduce or prevent later problems in school (National Research Council, 1998; Meisels and Margolis, 1988). But there is also a long and unhappy history with the unsophisticated use of IQ and achievement tests.

In a study of several hundred psychologists who work with young children, Bagnato and Neisworth (1994) found that only 4 percent of their respondents supported the use of norm-referenced, standardized intelligence tests for young children with developmental problems. Most respondents to their survey emphasized the importance of flexibility in the choice of assessment methods, the potential for modification of the instruments, and the need for a multidimensional, team-based assessment approach.

Potential problems with the use of norm-referenced tests are numerous. Some pertain to the technical adequacy of the instruments, and others derive from the way they are used (National Research Council, 1982b; Fuchs et al., 1987; Barnett and MacMann, 1992; National Research Council, 1997). In the first category are inadequate or unknown psychometric properties, including the common absence of children with disabilities in the samples used to develop test norms. Some children will require accommodations, but determining what accommodations are appropriate for whom and under what circumstances is difficult. The lack of knowledge about the functional characteristics of disability makes it difficult to determine whether or not the disability is related to the construct being measured, which in turn makes the interpretation of test results difficult. But even when tests are carefully developed, the test content may be inappropriate for certain subgroups of the population or biased against economically disadvantaged children. And finally, as is true of any assessment,

whether norm-referenced or not, the assessment may be irrelevant to the intervention process.

In addition to recognizing these challenges and problems, it is important to consider the assessment approach chosen in light of the purpose one has in mind. There are attributes of standardized, norm-referenced instruments that may make them well suited to certain high-stakes decisions, such as school accountability (although the issue of age appropriateness still obtains in preschool settings), but they should not be the cornerstone of an assessment system for working with individual children to help them develop new intellectual capacities, in which careful observation of the child in context is essential (Greenspan and Wieder, 1998). As Meltzer and Reid (1994) point out, standardized tests emphasize the end product of learning, ignoring the processes and strategies children use for problem solving. They fail to distinguish between a child's current level of performance and his or her ability to learn and acquire new skills and information. And they tend to ignore the role of motivation, personality, social factors, and cultural issues. As a consequence, the use of norm-referenced instruments has often led to misclassification and incorrect special education placements.

Perhaps the most important thing to remember about assessment of young children—whether the children are disabled, high risk, or developing typically—is that their development is episodic and uneven, with great variability within and among children. Intelligence, however one defines it, is not a stable construct in young children (e.g., Cronbach, 1990; Anatsi, 1988). This is manifest in the lack of agreement across measures and in the unreliability of assessment instruments. Standardized, norm-referenced tests are particularly vulnerable to misinterpretation because they imply a degree of certainty that assessments of young children simply cannot provide.

All of these cautionary statements about developmental and cultural issues and the potential shortcomings and misuses of standardized tests do not alter the fact that assessment is a key ingredient in the teaching and learning process. Assessment, whether of the informal variety that nearly all teachers engage in on a spontaneous basis, or of a more formal kind, can help to guide instruction and is an integral part of learning.

ASSESSMENT FOR PEDAGOGICAL AND INSTRUCTIONAL PLANNING

We know from research that learning is a process of building new understandings on the foundation of existing understandings. Learning will be most effective, therefore, when the child's preconceptions are engaged. This has direct implications for teaching and for the development of curricula. It is essential for teachers to ascertain the nature of thinking and the extent of learning for each child in order to make good decisions about what concepts, materials, and learning experiences will support the child's further growth. Perhaps the most significant change to take place in early childhood assessment in recent years concerns the linking of assessment and instruction. In their report to the National Education Goals Panel, Shepard and colleagues (1998) put it succinctly: "Assessing and teaching are inseparable processes."

The idea behind the fusion of assessment and instruction is relatively simple and rests on three fundamental assumptions (Meisels and Atkins-Burnett, 2000). The first is that assessment is a dynamic enterprise that calls on information from multiple sources collected over numerous time points, reflecting a wide range of child experiences and caregiver interpretations. The second assumption is that the formal act of assessment is only the first step in the process of acquiring information about the child and the family. Through intervention—by putting into practice the ideas or hypotheses raised by the initial assessment procedures—more information will be acquired that can serve the dual purpose of refining the assessment and enhancing the intervention. Third, assessment is of limited value in the absence of instruction or intervention. The meaning of an assessment is closely tied to its utility—to its contributions to decision making about practice or intervention or its confirmation of a child's continuing progress.

Nearly all early childhood educators rely on some form of informal monitoring of child learning in order to design programs and plan curricula—that is, in order to engage in pedagogy. However, relatively few early childhood teachers systematically observe, record, evaluate, and document children's learning—al-

though the need for systematic documentation is rapidly being imposed from many directions, including the new Head Start child performance standards.

Teachers can learn to observe and document children's skills, knowledge, and accomplishments as they participate in classroom activities and routines, interact with peers, and work with educational materials. Curriculum-embedded forms of assessment, for example, are contextualized methods that allow children opportunities to demonstrate their knowledge or skills through active engagement in classroom activities. Teachers who practice curriculum-embedded assessment rely on checklists, portfolios, and other collections of children's work to document learning and to monitor instruction (Meisels, 1996a, 1987; Wiggins, 1998).

Assessment of Competencies in Young Children

Research on cognition shows that young children's knowledge is more complex than expected. We are a long way from being able to integrate knowledge of developing competence and assessment methodology and practice. At present, a National Research Council Committee on the Foundations of Assessment, chaired by Robert Glaser and James Pellegrino, is attempting to rethink and enrich methods of assessment in light of advances in the science of human cognition, development, and learning. The goals they hold out for a new science of assessment are maximizing student competence, making students' reasoning processes more transparent, and creating a system where curriculum, instruction, and assessment are integrated in a mutually supportive fashion.

As learning scientists, measurement experts, and practitioners gradually create a new science and practice of assessment, there are several useful assessment methods that can be used to help to dig beneath the surface of overt behavior to get at thought processes. Chapter 5 described the use of specially designed tasks that can help uncover a child's grasp of important concepts, for example, the idea of "quantity" that is fundamental to understanding mathematics or the idea of "representation" that underlies words and illustrations that is the foundation of literacy. We

look below in detail at two approaches to uncovering such information: clinical interview methods (Ginsburg, 1997), and dynamic assessment (Bodrova and Leong, 1996; Burns, 1996; Burns et al., 1992; Day et al., 1997; Lidz and Pena, 1996), which often employs combinations of special tasks, observation, and clinical interviews. The discussion of assessing competence then moves on to performance assessment which, in contrast, focuses on concrete, observable behavior.

Clinical Interview

Piaget developed the “clinical interview” method, which is a most informative—and difficult—technique by which to assess children’s thinking. While the technique is most commonly used by trained therapists to assess children with learning problems and other disabilities, the clinical interview also lends itself to use by the teacher when the knowledge she wants to gain about a child is not evident in his or her performance.

The goal of the clinical interview is to identify the child’s underlying processes of thought. Its essence is its flexible, responsive, and open-ended nature. In the clinical interview, the interviewer asks the child to reflect on and articulate thinking processes. Although at the outset the interviewer has available several tasks likely to be appropriate for the topic at hand, initial questions are intentionally quite general, allowing the child’s response to influence the direction and content of the interview. The interview is a highly theory-bound activity that employs non-specific questions such as “How did you do it?” “What did you say to yourself?” and “How would you explain it to a friend?” so as to encourage rich verbalization and to avoid biasing the response. As the interview evolves, tasks and questions are determined in part by the child’s responses. Tasks are varied and modified, becoming more specific in order to focus on particular aspects of thinking, and more difficult in order to test the limits of understanding. In the clinical interview, the examiner’s behavior is to some degree contingent on the child’s; in standardized testing, the child’s behavior is always contingent on the examiner’s questions.

The clinical interview permits the interviewer to formulate,

test, and revise hypotheses about the child's thinking. The interview combines several methods: observation, test, experimentation, and "think aloud." The interviewer observes the child's behavior and listens to the child's verbalizations; presents "test items"—problems of various sorts, often involving concrete objects; experiments with different questions or tasks to test hypotheses; and asks the child to think aloud, to verbalize thought processes as explicitly as possible.

An interview is time-consuming and demanding, requiring 20 minutes to an hour of concentrated effort. A good interviewer must have command of the relevant content being assessed and must be familiar with typical thinking—that is, with normative behavior—at the child's level. Because it is a highly interactive technique, the interviewer must be able to generate useful hypotheses concerning the child's thinking on the spot, must have the ability to devise methods for testing these hypotheses as the interview proceeds, and must be sensitive to the nuances of the child's affect and motivation so as to establish rapport and motivation.

Many would agree that the interview method is powerful. But can it be used effectively by ordinary teachers? Or more precisely, can teachers who put the effort into learning the method make practical use of it in the hurly-burly of the everyday classroom? Several different approaches to adapting the interview method for classroom use have been suggested.

The National Council of Teachers of Mathematics (NCTM), for example, has been advocating the use of "authentic" assessment in classrooms, including the conduct of informal interviews. The NCTM journals frequently describe interview methods for teachers and give examples of their use. An article by Jencks (1989) described a procedure for interviewing students every five to six weeks at the beginning of a new topic of study. Another study reported research involving administration of 10- to 15-minute interviews by a mathematics specialist, who was then able to uncover difficulties hidden by correct responses on tests and to provide diagnostic information helpful to the teacher and the students' parents (Dionne and Fitzback-Labrecque, 1989). Others have described how interview activities can be integrated into, and can indeed transform, classroom instruction (Ginsburg et al.,

1993; Moon and Schulman, 1995). The experience with classroom use of interview methods is limited and the opportunities for adequate teacher training even more so. Yet there is much to recommend the approach to the research, practitioner, and professional development communities.

For one category of children, namely children with disabilities, the clinical interview is a critically important approach to assessment and will usually need to involve a trained therapist working with teachers. Most children with disabilities have problems in more than one area, which makes it very easy to make incorrect assumptions about the child's capabilities. Many children diagnosed with cognitive deficits, for example, also have problems with sensory processing and motor planning. It is essential to understand the unique profile of a child with special needs, to observe how that child interacts with family, teachers, and caregivers, in order to create the optimal intervention program tailored to the child's specific needs (Greenspan and Wieder, 1998:22-23).

Assessment in the Vygotskian Mode

One essential aspect of assessment to support learning is to provide information concerning children's ability to profit from teaching. The teacher is not interested in what the child knows or has mastered at any given point in time for its own sake, but as a clue to what concepts, knowledge, and opportunities can be provided in order to extend the child's emergent understandings. The goal is to understand a child's zone of proximal development—that area where learning is within reach but takes the child just beyond his or her existing ability (see Chapters 2 and 5). From this perspective, the role of assessment is to provide insight into the kind of educational experiences that will be most effective in helping particular children learn (Bodrova and Leong, 1996; Burns, 1996; Burns et al., 1992; Cronbach, 1990; Day et al., 1997; Ginsburg et al., 1999; Lidz and Pena, 1996)

In recent years, the concepts of two major theorists, Lev Vygotsky and Reuven Feuerstein, have stimulated and legitimated efforts to develop assessment techniques designed to promote children's learning potential. Vygotsky's theory describes

the human being as goal directed, an active seeker of information. Children come to formal education with a range of prior skills, understandings, knowledge, beliefs, and concepts built on experience which help the child navigate the surrounding world. These prior conceptions influence what the child notices about the environment and how they interpret it (National Research Council, 1999b, 1999c). The role of assessment, from this point of view, is to draw out and make explicit the child's prior conceptions or skills so that the teacher knows how and where to intervene to help the child advance. What the child is capable of at the present time becomes the pedagogical bridge to what a child can do, given assistance.

This pedagogical framework encourages the fusion of instruction and assessment. Consider the example of an assessment of children's equilibrium. Bodrova and Leong (1996) describe Teresa and Linda as they walk a balance beam: ". . . neither Teresa nor Linda can walk across a balance beam. Both of them stand on the end and stare down the beam. The teacher holds out her hand to assist each girl's performance. Although each is given the same teacher support, Teresa can only stand on the balance beam holding the teacher's hand tightly while Linda walks across the beam easily. Independent performance is misleading in this example. When we see how the two girls respond to assistance, we can tell that they are at very different levels." If the question motivating assessment were "Can this child walk the balance beam?" (or, Can this child add and subtract?), then the test would have stopped with each child's standing immobilized at the end of the balance beam and the simple answer for both would be "No." But because the emphasis is on understanding each child's current level of functioning (each child's zone of proximal development) as a guide for instruction, the more flexible and interactive assessment provided teachers with important information about what each particular child would need.

Similarly, Feuerstein (1979), much of whose work centered around the assessment of disadvantaged children's mental abilities, proposed a system of "dynamic assessment," in which the examiner engages in assisted instruction as a method for measuring the child's learning potential. Dynamic assessment techniques have also been designed to measure one or more skills

accurately and meaningfully over time. Teachers can then use repeated measurement on those behaviors to (a) model growth, (b) describe student difficulties, and (c) identify and plan programs for children who warrant intervention early in their lives. The disadvantage of this method is that the long-term indicators are only as good as the initial measurement—if an inaccurate measurement is used, the models developed will be inaccurate.

Dynamic Indicators of Basic Early Literacy Skills (DIBELS; Good and Kaminski, 1996; Kaminski and Good, 1998) is an example of a dynamic indicator. DIBELS focuses on story retelling, picture description, picture naming, letter naming, letter sounds, rhyming fluency, blending fluency, phonemic segmentation fluency, and onset recognition fluency tasks—the behaviors thought to represent the “critical” prereading skills needed for entering and succeeding in first grade. Important progress toward the development of a system to model growth, describe student difficulties, and identify and plan programs for children who warrant intervention early in their lives has been made; DIBELS’s phonemic segmentation fluency measure demonstrates strong traditional reliability and validity and suggests promise in terms of its capacity to model student literacy growth. Additional research on DIBELS and other alternative systems to monitor children’s development is needed to define their relative strengths and weaknesses.

Performance Assessment

Performance assessment takes a somewhat different approach to the assessment of competence. It is best understood in the context of learning about children’s knowledge, skills, and accomplishments through observing, recording, and evaluating their performance or work. Many feel that performance assessments lessen the likelihood of invidious comparisons between children, since each is evaluated according to how his or her specific levels of performance conform to the aims of the curriculum, rather than on how closely the performance conforms to the average performance of a normative group. In addition, they are not typically designed to sort and categorize children.

Some performance assessments can be described as “authen-

tic assessment” when they avoid “on-demand” tasks and focus instead on the assessment of concrete, observable behaviors on real (or realistic) tasks that are part of children’s ordinary classroom experiences. To the epistemological question “What is knowing?” performance assessment answers that the evidence of knowing is in the doing (see Meisels et al., 1995a). Hence, authentic performance assessments thrive on context and on the evidence acquired from natural settings.

With this focus on the evidence of knowing as represented in concrete behaviors or products, competence is not assessed on the basis of a single performance. Performance assessments require multiple sources of information and multiple observations of the same or related phenomena before conclusions can be drawn. They rely on extensive sampling of behavior in order to derive meaningful conclusions about individual children. A variety of documentation methods (e.g., a portfolio, a set of systematic checklists) can be brought into the assessment system. Over time and in the context of numerous performances, teachers observe “the *patterns* of success and failure and the reasons behind them” (Wiggins, 1998:705). These patterns constitute the evidence on which the assessment is based.

A significant virtue of performance assessments is that they permit children to demonstrate different approaches to performance. Different children may have highly comparable skills, but they may demonstrate these skills in very different ways. (Examples of performance assessment are presented in Box 6-1). Many also believe that a classroom emphasis on hands-on performance can enhance children’s motivation and offer a more informative way of engaging families in their children’s intervention progress.

There are several characteristics common to performance assessment that make the technique particularly attractive to many who work in the early childhood field (see Calfee, 1992; Herman et al., 1992; Shepard, 1991; Wiggins, 1998). They can encourage systematic processes of:

- documenting children’s daily activities to show their initiative and creativity,

- providing an integrated means for evaluating the quality of children's performance and behavior,
- reflecting on an individualized approach to pedagogy,
- evaluating those elements of learning and development that most conventional assessments do not capture very well,
- utilizing the information acquired in the teaching process to further elaborate the evaluative picture of the child that is emerging from the assessment, and
- shifting the teacher's attention and activity away from the typical content of test taking and onto the learning of the child and the environment in which instruction is taking place.

Traditional norm-referenced ability and achievement tests provide a summative statement about the test taker. An important point to be made about all three of the approaches to the assessment of competence discussed here, including performance assessment, is that they are formative: they provide information that can be used both to change the process of intervention and to keep track of children's progress and accomplishments. Information about the child and the setting that is gathered on a structured but continuing basis is then used to inform the intervention-instructional process. Because the emphasis is on continuous assessment, they can be used to monitor a child's progress frequently, rather than summarizing that progress on annual or semiannual occasions. But performance assessments, like all of the assessments, will only be as strong as the theory on which they are based. Assessment involves theorizing—having informed ideas about the processes of learning and developing hypotheses about a child's strengths and deficits on the basis of assessment information.

Instructional Assessment and Pedagogy

When pedagogy is defined as it has been in this volume—as an interactional construct that reflects a joint focus on the child's status and the characteristics of the educational setting—two conditions are critical for the assessment of learning (see Meisels, 1999, for an elaboration of these ideas). First, there must be sus-

BOX 6-1 **Approaches to Performance Assessment**

Work Sampling System

Widely used throughout the nation since 1991, the Work Sampling System (Meisels, 1987; Meisels et al., 1994) is a performance assessment designed for children from preschool through grade 5. This approach relies on developmental guidelines and checklists, portfolios, and summary reports. It is based on using teachers' perceptions of their students in actual classroom situations while simultaneously informing, expanding, and structuring those perceptions. It involves students and parents in the learning and assessment process, instead of relying on measures that are external to the community, classroom, and family context, and it makes possible a systematic documentation of what children are learning and how teachers are teaching.

The Work Sampling System draws attention to what the child brings to the learning situation and what the learning situation brings to the child. As active constructors of knowledge, children should be expected to analyze, synthesize, evaluate, and interpret facts and ideas. This approach to performance assessment allows teachers the opportunity to learn about these processes by documenting children's interactions with materials, adults, and peers in the classroom environment and using this documentation to evaluate children's achievements and plan future educational interventions. Evidence of the reliability and validity of the Work Sampling System with kindergarten children is available (Meisels et al., 1995b; Meisels et al., 1998).

Child Observation Record

Developed by High/Scope, this assessment provides a means of systematically observing children's activities in the ongoing con-

tained opportunities for the interactions between teacher and child to occur, and, second, these interactions must occur over time, rather than on a single occasion. This view does not hold that one can round up all of the kindergarten children in a community on a given day and test them to determine what they know and can do. Rather, it suggests that learning can be assessed only over time and in context.

Several methods exist today that can provide the type of as-

text of their classroom experiences, including prolonged activity and across time periods (High/Scope Research Foundation, 1992). The focus of the observations is on “important developmental experiences that should happen in all developmentally appropriate early childhood programs” and on “existing strengths and weaknesses rather than skills that have not yet emerged.” Six broad areas are assessed: initiative, social relations, creative representation, music and movement, language and literacy, and logic and mathematics. The system is comprehensive, providing behaviors to observe, a systematic way to collect anecdotal remarks, and a means to draw conclusions about the children’s performance in order to plan instruction.

Project Construct

“Project Construct is a process-oriented curriculum and assessment framework for working with children ages three through seven” (Missouri Department of Elementary and Secondary Education, 1992:3). It is based on constructivist theory and includes curriculum and assessment guidelines organized into four interrelated domains: sociomoral, cognitive, representational, and physical development. The project design provides a variety of resources for educators and parents, including curriculum materials, assessment instruments, and training and professional development opportunities. The Project Construct Assessment System is an integrated set of evaluation tools aligned with the Project Construct curriculum goals for children. Two components make up the assessment system—the Formative Assessment Program and the Inventory. Both parts utilize multiple sources of information that are primarily collected by teachers over extended periods of time.

assessment that occurs over time and in interaction. They contain not only a joint focus on the child’s status and the characteristics of the child’s educational setting, but they also encourage individual planning, programming, and evaluation. The Work Sampling System, designed for preschool-grade 5 (Meisels et al., 1994; Meisels, 1996b), is one example of an assessment system designed to achieve these goals.

The three types of instructional assessment described above

are not adopted easily or without expense. They require extensive professional development for teachers; changes in orientation regarding testing, grading, and student classification by educational policy makers; and alteration in expectations by parents and the community. Such changes entail financial burdens, centralized coordination and program evaluation, and long-term commitment from teachers, parents, and the community—all of which are potential obstacles to implementation.

The path to progress has been demarked in the most recent call to the field from the Goal 1 Technical Planning Group of the National Education Goals Panel: “The Technical Planning Group, while understanding the complexity of the technical challenges associated with defining and assessing early development and learning . . . is convinced that new assessments are doomed to repeat past problems unless such efforts are permeated by a conceptual orientation that accommodates cultural and contextual variability in *what* is being measured and in *how* measurements are constructed. Within the broad parameters of standardization, then, flexibility and inventiveness must be brought to bear on the content and the process of assessment” (Kagan et al., 1995:42).

ASSESSMENT FOR SELECTION AND DIAGNOSIS

Two important functions served by testing are selection and diagnosis. Selection or “readiness” assessments are intended to determine a child’s preparedness to profit from a particular curriculum. Diagnostic testing is used to determine the type and extent of a special need or disability. A third type of assessment, developmental screening, is a relatively brief testing instrument typically used to determine whether further diagnostic testing is indicated. Each type of assessment is quite distinct from the others.

Developmental Screening

Developmental screening is a brief procedure designed to identify children at high risk for school failure. These are norm-referenced standardized tests that typically evaluate a broad range of abilities, including intellectual, emotional, social, and

motor abilities. Developmental screening is typically performed individually on large numbers of children, requiring very little time per child. There are several instruments that have been developed with attention to *Standards for Educational and Psychological Testing* (APA, AERA, NCME, 1986) (AERA, APA, NCME 2000) and have high reliability and predictive validity (Meisels, 1987, 1988; Nuttall et al., 1999). When appropriate instruments are used, developmental screening is an extremely valuable source of information. They should be considered as the first step in an evaluation and intervention process that can help prevent the emergence of more serious problems in children before they have had an opportunity to affect the course of development (Meisels and Atkins-Burnett, 1994).

Diagnostic Assessment

Diagnostic assessment is intended to determine conclusively whether a child has special needs, ascertain the nature and character of the child's problems, and suggest the cause of the problems, if possible. According to federal law, diagnostic assessments administered by schools must be conducted in a team setting that utilizes multiple sources of data and is part of a system of special education services. Such an assessment provides the data that are used to create individualized family service plans (IFSPs) and individual educational plans (IEPs) (Bailey and Wolery, 1992).

In the past, the most common tools used for diagnostic purposes were intelligence tests, which focused primarily on the child in isolation. Today, assessments are considered incomplete unless they view the child in relation to three domains. The first is the child's biology. The second is the child's interactive patterns with parents, teachers, siblings, and others. And the third is comprised of the patterns of the family, the culture, and the larger environment (Greenspan and Wieder, 1998; Greenspan, 1992). Unless the child is examined within these contexts, inferences about his or her developmental status will be incomplete and generalizations about developmental trajectories may be seriously flawed.

Greenspan and Wieder (1998) posit six fundamental develop-

ment skills that lay the foundation for all learning and underlie all advanced thinking, problem solving, and coping (pp. 3-4):

1. The dual ability to take an interest in the sights, sounds, and sensations of the world and to calm oneself down.
2. The ability to engage in relationships with other people.
3. The ability to engage in two-way communications.
4. The ability to create complex gestures, to string together a series of actions into an elaborate and deliberate problem-solving sequence.
5. The ability to create ideas.
6. The ability to build bridges between ideas to make them reality-based and logical.

These “functional emotional skills” provide the theoretical framework for assessing the child’s developmental progress over time and guide the course of interventions.

Greenspan draws heavily on the clinical interview procedures described earlier as a means of getting underneath the disability categories that so influence our expectations of children (autism, attention deficit disorder, mental retardation, pervasive developmental disorder) to the “functional emotional skills” of the individual child. He argues compellingly from his work with infants and young children that the differences among children who bear the same label are greater than their similarities and that, with careful assessment, it is possible to tailor a treatment approach that helps the individual child climb the developmental ladder.

Whatever combination of assessments is used for the purposes of diagnosing disabilities and learning problems, it is important that any cognitive, behavioral, or sensory measures used meet high standards of validity and reliability. The very real challenges of interpreting the performance of children with special needs on standardized instruments also means that it should be in the hands of trained professionals.

Readiness Testing

Readiness tests indicate a child’s relative preparedness to participate in a particular classroom, rather than addressing general

developmental status. The most commonly used readiness tests include items that assess children's perceptual skills (matching one shape from an array of other shapes), knowledge of alphabet letters, awareness of the use of prepositions (on, under, behind), colors, and sometimes receptive vocabulary. Note that the emphasis is very different from Greenspan's functional emotional skills listed above.

Early learning readiness measures are widely used—many would say misused—to determine whether children are ready for kindergarten. For example, many programs designed to prepare poor and immigrant children for kindergarten or first grade use readiness tests during the last year of preschool, typically at 5 years of age, to make promotion recommendations.

It is interesting to note that what readiness tests measure is not well aligned with what teachers think is important. Teachers' views of readiness were surveyed by the U.S. Department of Education's Kindergarten Teacher Survey on Student Readiness (National Center for Education Statistics, 1999). As shown in Figure 6-1, over 75 percent of the teachers surveyed considered it very important or essential that children be physically healthy, rested, nourished, enthusiastic and curious in approaching new activities, and able to communicate needs, wants, and thoughts verbally in their primary language. In contrast, 25 percent or fewer of the teachers considered the following items very important or essential: counts to 20, has good problem-solving skills, can use a pencil or paintbrush, and knows letters of the alphabet. In short, teachers' opinions about readiness seem to reflect the importance of receptivity to learning, rather than the particular skills that a child may or may not have acquired before coming to school.

While it is easy to endorse what the survey indicates that teachers think, it is also important to recognize that these characteristics fall far short of what the cognitive and developmental research shows that young children are capable of. This misalignment between current goals and future possibilities will eventually find a measure of resolution as advanced learning principles are incorporated into learning and instruction in preschool and child care settings and as more and more children have the advantage of such instruction.

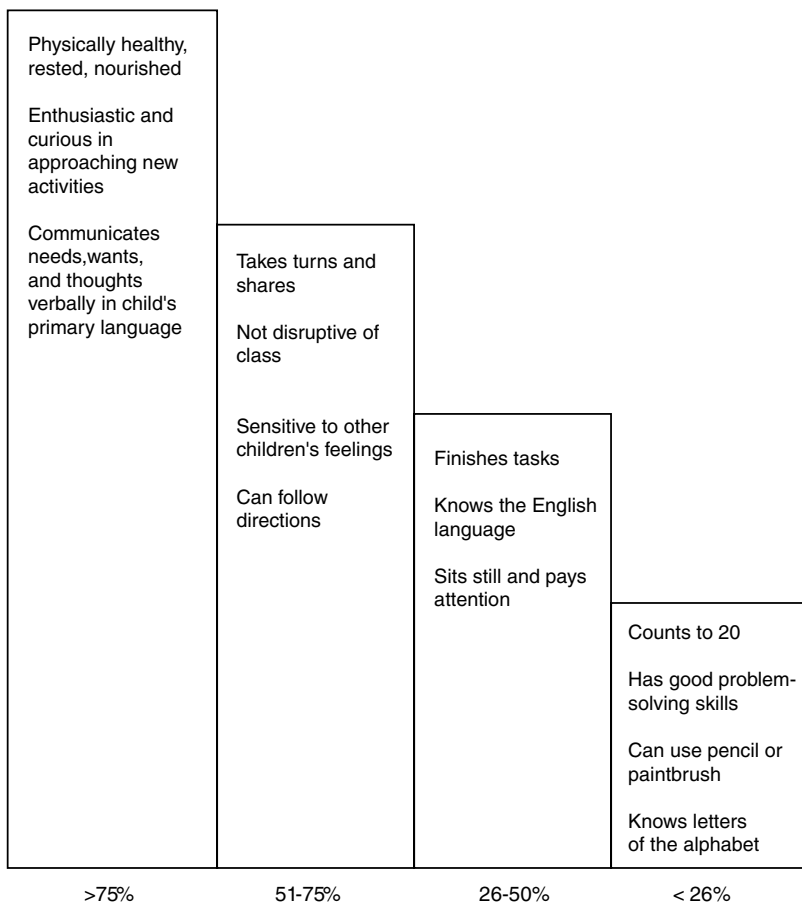


FIGURE 6-1 Percentage of public school kindergarten teachers indicating whether various factors for kindergarten readiness were very important or essential. SOURCE: National Center for Educational Statistics (1999).

In the meantime, schools and programs need to think carefully about the use of readiness tests. Readiness is a very complex construct, including intellectual and social abilities, and its assessment will be affected greatly by young children's episodic and unstable growth patterns and by variations in how children live and are raised. Some children may do very poorly on readiness tests at the outset of school simply because they were not exposed to or taught the items that are on tests. Once enrolled in

kindergarten, these same children may thrive. This is a particular concern for children from minority and disadvantaged backgrounds. And since schools and programs differ, the fundamental requirement in every evaluation of a child's school readiness should be that the assessment is grounded in direct relevance to the criterion, namely, functioning in that school or program.

These considerations have led many to conclude that readiness tests are not suited for use in child placement and promotion decisions, although they may have value for purposes of instructional planning (Meisels, 1987, 1989a, 1989b; Stallman and Pearson, 1990).

ASSESSMENT FOR POLICY DECISIONS

Tests and assessment results are increasingly used as a basis for important policy decisions in education. Large-scale testing programs generate data that inform about which schools or programs should be funded, which should be closed, who should be rewarded, what types of programs should be developed, and who should be informed that improvement is required if further assistance is to be forthcoming. Public reporting of assessment data by district or by school has become commonplace, as has the use of these data for rewards and potential sanctions. These types of decisions are known as high-stakes decisions (see Madaus, 1988; National Research Council, 1997).

High-stakes testing also refers to the use of assessment data to make decisions about individual students or teachers. The use of readiness tests to make decisions about enrolling a child in kindergarten provides one illustration. Other uses include retention, promotion, tracking, placement in special education, and selection into advanced programs (Madaus, 1988; Meisels, 1989a, 1989b; National Research Council, 1999a).

High-stakes testing is closely tied to the notion of accountability, so that poor scores on such examinations will result in negative sanctions of one sort or another. It is widely believed that tangible rewards or punishments will provide strong incentives for schools, teachers, and children to improve their performance.

The use of assessment to support policy decisions is much

more prevalent in the public school system than in preschool settings. However, most early childhood programs for poor children and children with disabilities are supported with public monies, and agencies typically seek to evaluate the effectiveness of their programs in order to justify the spending of taxpayers' dollars. The evaluation studies of Head Start and other such programs are of this genre. As more and more young children are cared for and educated outside the home, the pressure for accountability is likely to increase—not just to satisfy a demand for reporting on public expenditures but as an expression of society's interest in protecting its youngest and most vulnerable members.

Such uses of assessment data for purposes external to the classroom, rather than improve educational practice directly, place a particularly heavy burden both on the assessment instruments and on the responsible adults. The data must be collected in a standardized way that permits comparisons across schools. This means, for example, that teachers should not give help during the assessment unless it is part of the standard administration to do so (Shepard et al., 1998). Note how different the protocols for appropriate test use are for this sort of testing than for assessment designed to support pedagogy and instruction.

Although there are many attempts under way to rethink testing and assessment to combine some of the statistical power and generalizability of standardized tests with the richer portrait of individual learning and development that characterizes many alternative assessments, we are a long way from achieving that goal. Some researchers are proposing that the aggregate of classroom-based assessment be used for accountability reporting (Bridgeman et al., 1995). If the use of external standardized tests increases in the preschool environment for reasons of public policy, it is essential that they meet the highest standards of reliability and validity.

Above all, any tests used for policy purposes must not be mistaken for statements about the learning trajectory of the individual child or allowed to diminish the importance of the kinds of assessments that will support learning. Likewise, standards for child performance such as those articulated by Head Start should absolutely not have any consequences associated with

them for the individual child (AERA, APA, NCME, 2000; Shepard et al., 1998).

SUMMARY

What are the roles of assessment in preschool? We have described three broad categories that constitute the major purposes of assessment in early childhood settings—assessment to inform instruction, assessment for diagnostic and selection purposes, and assessment for accountability and program evaluation. Just as there are different purposes for assessment, there are many different types of assessments, from the clinical interview to the statewide assessment used for school accountability. No single assessment will satisfy all educational needs or solve all educational problems.

Assessments must be used carefully and appropriately if they are to resolve, and not create, educational problems. This means using each assessment in the way in which it was designed and intended. To use assessment as a blunt instrument, in which one type of assessment is expected to perform the functions of others, squanders resources and places children at risk for school difficulties. The Committee on School Health and the Committee on Early Childhood of the American Academy of Pediatrics (1995) made clear the dangers inherent in the inappropriate use of tests and assessments (p. 437):

When instruments and procedures designed for screening are used for diagnostic purposes, or when tests are administered by individuals who have a limited perspective on the variations of normal development, or when staff with little formal training in test administration perform the screening, children can be wrongly identified and their education jeopardized.

We have written at length about the need for a fusion of assessment and instruction in early childhood settings. Assessment has an important role to play in revealing a child's prior knowledge, development of concepts, and ways of interacting with and understanding the world so that teachers can choose a pedagogical approach and curricular materials that will support the child's further learning and development. We have described a number

of promising approaches to assessment to support learning—the clinical interview, dynamic assessment, performance assessment.

The fact is, however, that most early childhood educators are not trained in traditional testing and measurement, to say nothing of the newer kinds of assessment. Moreover, assessment in early childhood tends to be considered external and irrelevant to the teaching and learning process, rather than something that can complement educational programs and, indeed, is essential to making the program work for each child. If we are to use the important findings about human learning from the cognitive, neurological, and developmental sciences to improve early childhood pedagogy and instruction, it is important that early childhood educators and caregivers be trained to use assessments for purposes that will advance teaching and learning (Arter, 1999; Brookhart, 1999; Jones and Chittenden, 1995; Meisels, 1999; Sheingold et al., 1995; Stiggins, 1991, 1999).

Finally, we have emphasized the importance of using assessments and tests particularly carefully with young children. The first five years of human life are a time of incredible growth and learning. The rapid growth of the brain in the early years provides an opportunity for the environment to play an enormous role in development. But the course of development in young children is uneven and episodic, with great spurts in learning in one and lags in another. As a consequence, assessment results can easily be misinterpreted. Standardized tests are particularly vulnerable to misuse with this population, but any assessment procedures must be used intelligently and with care. The developmental characteristics of young children make it even more important that teachers and caregivers be trained to think about and use assessment well.