

CHAPTER 1

An Overview of Assessment in Early Childhood



Chapter Objectives

As a result of reading this chapter, you will be able to

1. Understand the purposes of assessment in early childhood
2. Understand different meanings of the term *assessment*
3. Understand the history of tests and measurements in early childhood
4. Develop an awareness of issues in testing young children

Understanding Assessment in Infancy and Early Childhood

Not too long ago, resources on early childhood assessment were limited to occasional articles in journals, chapters in textbooks on teaching in early childhood programs, and a few small textbooks that were used as secondary texts in an early childhood education course. Very few teacher preparation programs offered a course devoted to assessment in early childhood. Now, in the 21st century, assessment of very young children has experienced a period of very rapid growth and expansion. In fact, it has been described as a “virtual explosion of testing in public schools” (Meisels & Atkins-Burnett, 2005, p. 1).

There has also been an explosion in the numbers of infants, toddlers, and preschoolers in early childhood programs and the types of programs that serve them. Moreover, the diversity among these young children increases each year. Currently, Head Start programs serve children and families who speak at least 140 different languages. In some Head Start classrooms, ten different languages might be used. Head Start teaching teams may also be multilingual, also representing diversity (David, 2005).

What Is Assessment?

What do we need to know about all these diverse children with all kinds of families, cultures, and languages? The study of individuals for measurement purposes begins before birth with assessment of fetal growth and development. At birth and throughout infancy and early childhood, various methods of measurement are used to evaluate the child’s growth and development. Before a young child enters a preschool program, he or she is measured through medical examinations. Children are also measured through *observations* of developmental milestones, such as saying the first word or walking independently, by parents and other family members. Children might also be screened or evaluated for an early childhood program or service. Assessment is really a *process*. A current definition describes the assessment process: “Assessment is the process of gathering information about children from several forms of evidence, then organizing and interpreting that information” (McAfee, Leong, & Bodrova, 2004, p. 3).

Assessment of children from birth through the preschool years is different from assessment of older people. Not only can young children not write or read, but also the young developing child presents different challenges that influence the choice of measurement strategy, or how to measure or assess the child. Assessment methods must be matched with the level of mental, social, and physical development at each stage. Developmental change in young children is rapid, and there is a need to assess whether development is progressing normally. If development is not normal, the measurement and evaluation procedures used are important in making decisions regarding appropriate intervention services during infancy and the preschool years.

Purposes of Assessment

Assessment is used for various purposes. We may want to learn about individual children. We may conduct an *evaluation* to assess a young child's development in language or mathematics. When we need to learn more, we may assess the child by asking her or him to describe what she or he has achieved. For example, a first-grade teacher may use measurement techniques to determine what reading skills have been mastered and what weaknesses exist that indicate a need for additional instruction.

Assessment strategies may be used for *diagnosis*. Just as a medical doctor conducts a physical examination of a child to diagnose an illness, psychologists, teachers, and other adults who work with children can conduct an informal or formal assessment to diagnose a developmental delay or identify causes for poor performance in learning.

If medical problems, birth defects, or developmental delays in motor, language, cognitive, or social development are discovered during the early, critical periods of development, steps can be taken to correct, minimize, or remediate them before the child enters school. For many developmental deficits or differences, the earlier they are detected and the earlier intervention is planned, the more likely the child will be able to overcome them or compensate for them. For example, if a serious hearing deficit is identified early, the child can learn other methods of communicating and acquiring information.

Assessment of young children is also used for *placement*—to place them in infant or early childhood programs or to provide special services. To ensure that a child receives the best services, careful screening and more extensive testing may be conducted before selecting the combination of intervention programs and other services that will best serve the child.

Program planning is another purpose of assessment. After children have been identified and evaluated for an intervention program or service, assessment results can be used in planning the programs that will serve them. These programs, in turn, can be evaluated to determine their effectiveness.

Besides identifying and correcting developmental problems, assessment of very young children is conducted for other purposes. One purpose is *research*. Researchers study young children to better understand their behavior or to measure the appropriateness of the experiences that are provided for them.

The National Early Childhood Assessment Resource Group summarized the purposes for appropriate uses of assessment in the early childhood years as follows:

Purpose 1: Assessing to promote children's learning and development

Purpose 2: Identifying children for health and social services

Purpose 3: Monitoring trends and evaluating programs and services

Purpose 4: Assessing academic achievement to hold individual students, teachers, and schools accountable (Shepard, Kagan, Lynn, & Wurtz, 1998).

(See Figure 2-1.)

How were these assessment strategies developed? In the next section, I describe how certain movements or factors, especially during the past century, have affected the development of testing instruments, procedures, and other measurement techniques that are used with infants and young children.

Early Intervention for a Child with Hearing Impairment

Julio, who is 2 years old, was born prematurely. He did not have regular checkups during his first year, but his mother took him to a community clinic when he had a cold and fever at about 9 months of age. When the doctor noticed that Julio did not react to normal sounds in the examining room, she stood behind him and clapped her hands near each ear. Because Julio did not turn toward the clapping sounds, the doctor suspected that he had a hearing loss. She arranged for Julio to be examined by an audiologist at an eye, ear, nose, and throat clinic.

Julio was found to have a significant hearing loss in both ears. He was fitted with hearing aids and is attending a special program twice a week for children with hearing deficits. Therapists in the program are teaching Julio to speak. They are also teaching his mother how to make Julio aware of his surroundings and help him to develop a vocabulary. Had Julio not received intervention services at an early age, he might have entered school with severe cognitive and learning deficits that would have put him at a higher risk for failing to learn.

The Evolution of Assessment of Young Children

Interest in studying young children to understand their growth and development dates back to the initial recognition of childhood as a separate period in the life cycle. Johann Pestalozzi, a pioneer in developing educational programs specifically for children, wrote about the development of his 3½-year-old son in 1774 (Irwin & Bushnell, 1980). Early publications also reflected concern for the proper upbringing and education of young children. *Some Thoughts Concerning Education* by John Locke (1699), *Emile* (Rousseau, 1762/1911), and Frederick Froebel's *Education of Man* (1896) were influential in focusing attention on the characteristics and needs of children in the 18th and 19th centuries. Rousseau believed that human nature was essentially good and that education must allow that goodness to unfold. He stated that more attention should be given to studying the child so that education could be adapted to meet individual needs (Weber, 1984). The study of children, as advocated by Rousseau, did not begin until the late 19th and early 20th centuries.

Scientists throughout the world used observation to measure human behaviors. Ivan Pavlov proposed a theory of conditioning to change behaviors. Alfred Binet developed the concept of a normal mental age by studying memory, attention, and intelligence in children. Binet and Theophile Simon developed an intelligence scale to determine mental age that made it possible to differentiate the abilities of individual

children (Weber, 1984). American psychologists expanded these early efforts, developing instruments for various types of measurement.

The study and measurement of young children today has evolved from the child study movement, the development of standardized tests, Head Start and other federal programs first funded in the 1960s, and the passage of Public Law 94-142 (the Individuals with Disabilities Education Act) and Public Law 99-457 (an expansion of PL 94-142 to include infants). Currently, there is a movement toward more meaningful learning or authentic achievement and assessment (Newmann, 1996; Wiggins, 1993). At the same time, continuing progress is being made in identifying, diagnosing, and providing more appropriate intervention for infants and young children with disabilities (Meisels & Fenichel, 1996).

The Child Study Movement

G. Stanley Hall, Charles Darwin, and Lawrence Frank were leaders in the development of the child study movement that emerged at the beginning of the 20th century. Darwin, in suggesting that by studying the development of the infant one could glimpse the development of the human species, initiated the scientific study of the child (Kessen, 1965). Hall developed and extended methods of studying children. After he became president of Clark University in Worcester, Massachusetts, he established a major center for child study. Hall's students—John Dewey, Arnold Gesell, and Lewis Terman—all made major contributions to the study and measurement of children. Dewey advocated educational reform that affected the development of educational programs for young children. Gesell first described the behaviors that emerged in children at each chronological age. Terman became a leader in the development of mental tests (Irwin & Bushnell, 1980; Wortham, 2002).

Research in child rearing and child care was furthered by the establishment of the Laura Spelman Rockefeller Memorial child development grants. Under the leadership of Lawrence Frank, institutes for child development were funded by the Rockefeller grants at Columbia University Teacher's College (New York), the University of Minnesota, the University of California at Berkeley, Arnold Gesell's Clinic of Child Development at Yale University, the Iowa Child Welfare Station, and other locations.

With the establishment of child study at academic centers, preschool children could be observed in group settings, rather than as individuals in the home. With the development of laboratory schools and nursery schools in the home economics departments of colleges and universities, child study research could also include the family in broadening the understanding of child development. Researchers from many disciplines joined in an ongoing child study movement that originated strategies for observing and measuring development. The results of their research led to an abundant literature. Between the 1890s and the 1950s, hundreds of children were studied in academic settings throughout the United States (Weber, 1984). Thus, the child study movement has taught us to use observation and other strategies to assess the child. Investigators today continue to add new knowledge about child development and learning that aids parents, preschool teachers and staff members, and professionals in institutions and agencies that provide services to children and

families. In the last decade of the 20th century and in the 21st century, brain research has opened up a whole new perspective of the nature of cognitive development and the importance of the early years for optimum development and later learning (Begley, 1997; Shore, 1997). These new findings have caused early childhood educators to reflect on the factors that affect early development and the implications for programming for children in infancy and early childhood.

Standardized Tests

Standardized testing also began around 1900. When colleges and universities in the East sought applicants from other areas of the nation in the 1920s, they found the high school transcripts of these students difficult to evaluate. The *Scholastic Aptitude Test (SAT)* was established to permit fairer comparisons of applicants seeking admission (Cronbach, 1990).

As public schools expanded to offer 12 years of education, a similar phenomenon occurred. To determine the level and pace of instruction and the grouping of students without regard for socioeconomic class, objective tests were developed (Gardner, 1961). These tests grew out of the need to sort, select, or otherwise make decisions about both children and adults.

The first efforts to design tests were informal. When a psychologist, researcher, or physician needed a method to observe a behavior, he or she developed a procedure to meet those needs. The procedure was often adopted by others with the same needs. When many people wanted to use a particular measurement strategy or test, the developer prepared printed copies for sale. As the demand for tests grew, textbook publishers and firms specializing in test development and production also began to create and sell tests (Cronbach, 1990).

American psychologists built on the work of Binet and Simon in developing the intelligence measures described earlier. Binet's instrument, revised by Terman at Stanford University, came to be known as the *Stanford-Binet Intelligence Scale*. Other Americans, particularly educators, welcomed the opportunity to use precise measurements to evaluate learning. Edward Thorndike and his students designed measures to evaluate achievement in reading, mathematics, spelling, and language ability (Weber, 1984). Because of the work of Terman and Thorndike, testing soon became a science (Scherer, 1999). By 1918, more than 100 standardized tests had been designed to measure school achievement (Monroe, 1918).

After World War II, the demand for dependable and technically refined tests grew, and people of all ages came to be tested. As individuals and institutions selected and developed their own tests, the use of testing became more centralized. Statewide tests were administered in schools, and tests were increasingly used at the national level.

The expanded use of tests resulted in the establishment of giant corporations that could assemble the resources to develop, publish, score, and report the results of testing to a large clientele. Centralization improved the quality of tests and the establishment of standards for test design. As individual researchers and teams of psychologists continue to design instruments to meet current needs, the high quality of these newer tests can be attributed to the improvements and refinements made over the years and to the increased knowledge of test design and validation (Cronbach, 1990).

Head Start and the War on Poverty

Prior to the 1960s, medical doctors, psychologists, and other professionals serving children developed tests for use with preschool children. Developmental measures, IQ tests, and specialized tests to measure developmental deficits were generally used for noneducational purposes. Child study researchers tended to use observational or unobtrusive methods to study the individual child or groups of children. School-age children were tested to measure school achievement, but this type of test was rarely used with preschool children.

After the federal government decided to improve the academic performance of children from low-income homes and those from non-English-speaking backgrounds, test developers moved quickly to design new measurement and evaluation instruments for these preschool and school-age populations.

In the late 1950s, there was concern about the consistently low academic performance of children from poor homes. As researchers investigated the problem, national interest in improving education led to massive funding for many programs designed to reduce the disparity in achievement between poor and middle-class children. The major program that involved preschool children was Head Start. Models of early childhood programs ranging from highly structured academic, child-centered developmental to more traditional nursery school models were designed and implemented throughout the United States (White, 1973; Zigler & Valentine, 1979).

All programs funded by the federal government had to be evaluated for effectiveness. As a result, new measures were developed to assess individual progress and the programs' effectiveness (Laosa, 1982). The quality of these measures was uneven, as was comparative research designed to compare the overall effectiveness of Head Start. Nevertheless, the measures and strategies developed for use with Head Start projects added valuable resources for the assessment and evaluation of young children (Hoepfner, Stern, & Nummedal, 1971).

Other federally funded programs developed in the 1960s, such as bilingual programs, Title I, the Emergency School Aid Act, Follow Through, and Home Start, were similar in effect to Head Start. The need for measurement strategies and tests to evaluate these programs led to the improvement of existing tests and the development of new tests to evaluate their success accurately.

Legislation for Young Children With Disabilities

PL 94-142

Perhaps the most significant law affecting the measurement of children was Public Law (PL) 94-142, the Education for All Handicapped Children Act, passed in 1975. This law, later amended and renamed the Individuals with Disabilities Education Act (IDEA), guaranteed all children with disabilities the right to an appropriate education in a free public school and placement in the least restrictive learning environment. The law further required the use of nondiscriminatory testing and evaluation of these children (McCollum & Maude, 1993).

The implications of the law were far reaching. Testing, identification, and placement of students with mental retardation and those with other disabilities were difficult. Existing tests were no longer considered adequate for children with special

One Family's Experience with Head Start

Rosa is a graduate of the Head Start program. For 2 years, she participated in a class housed in James Brown School, a former inner-city school that had been closed and remodeled for other community services. Two Head Start classrooms were in the building, which was shared with several other community agencies serving low-income families. In addition to learning at James Brown School, Rosa went on many field trips, including trips to the zoo, the botanical garden, the public library, and a nearby McDonald's restaurant.

This year Rosa is a kindergarten student at West Oaks Elementary School with her older brothers, who also attended Head Start. Next year, Rosa's younger sister, Luisa, will begin the program. Luisa looks forward to Head Start. She has good memories of the things she observed Rosa doing in the Head Start classroom while visiting the school with her mother.

Luisa's parents are also happy that she will be attending the Head Start program. Luisa's older brothers are good students, which they attribute to the background they received in Head Start. From her work in kindergarten, it appears that Rosa will also do well when she enters first grade.

needs. Classroom teachers had to learn the techniques used to identify students with disabilities and determine how to meet their educational needs (Kaplan & Saccuzzo, 1989).

The law required that a team of teachers, parents, diagnosticians, school psychologists, medical personnel, and perhaps social workers or representatives of government agencies or institutions be used to identify and place students with disabilities. When appropriate, the child must also be included in the decision-making process. The team screens, tests, and develops an Individual Education Programme (IEP) for each child. Not all team members are involved in every step of the process, but they can influence the decisions made.

The term **mainstreaming** came to define the requirement that the child be placed in the least restrictive environment. This meant that as often as possible, the child would be placed with children developing normally, rather than in a segregated classroom for students in special education. How much mainstreaming was beneficial for the individual student? The question was difficult to answer. In addition, the ability of teachers to meet the needs of students with and without disabilities simultaneously in the same classroom is still debated. Nevertheless, classroom teachers were expected to develop and monitor the educational program prescribed for students with disabilities (Clark, 1976).

The identification and diagnosis of students with disabilities is the most complex aspect of PL 94-142. Many types of children need special education, including students with mental retardation, physical and visual disabilities, speech impairments, auditory disabilities, learning disabilities, and emotional disturbances, and

students who are gifted. Children may have a combination of disabilities. The identification and comprehensive testing of children to determine what types of disabilities they have and how best to educate them requires a vast array of assessment techniques and instruments. Teachers, school nurses, and other staff members can be involved in initial screening and referral, but the extensive testing used for diagnosis and prescription requires professionals who have been trained to administer psychological tests (Mehrens & Lehmann, 1991).

Under PL 94-142, all children with disabilities between ages 3 and 21 are entitled to free public education. This means that preschool programs must also be provided for children under age 6. Public schools have implemented early childhood programs for children with disabilities, and Head Start programs are required to include them (Guralnick, 1982; Spodek & Saracho, 1994). Other institutions and agencies also provide programs for children with and without disabilities.

PL 99-457

Many of the shortcomings of PL 94-142 were addressed in PL 99-457 (Education of the Handicapped Act Amendments), passed in 1986. The newer law authorized two new programs: the Federal Preschool Program and the Early Intervention Program. Under PL 94-142, the state could choose whether to provide services to children with disabilities between ages 3 and 5. Under PL 99-457, states must prove that they are meeting the needs of all these children if they wish to receive federal funds under PL 94-142. The Federal Preschool Program extends the right of children with disabilities under PL 94-142 to all children with disabilities between ages 3 and 5.

The Early Intervention Program established early intervention services for all children between birth and age 2 who are developmentally delayed. All participating states must now provide intervention services for all infants and toddlers with disabilities (McCollum & Maude, 1993; Meisels & Shonkoff, 1990).

How to measure and evaluate young children with disabilities and the programs that serve them are a continuing challenge (Cicchetti & Wagner, 1990). The design of measures to screen, identify, and place preschool children in intervention programs began with the passage of PL 94-142 and was extended under PL 99-457. Many of these instruments and strategies, particularly those dealing with developmental delay, were also used with preschool programs serving children developing normally, as well as those with developmental delays or disabilities.

As children with disabilities were served in a larger variety of settings, such as preschools, Head Start programs, child-care settings, infant intervention programs, and hospitals, early childhood educators from diverse backgrounds were involved in determining whether infants and young children were eligible for services for special needs. Early childhood educators and other practitioners in the field were challenged to be knowledgeable in measurement and evaluation strategies for effective identification, placement, and assessment of young children in integrated early childhood settings (Goodwin & Goodwin, 1993).

Many questions were raised about appropriately serving young children with diverse abilities. Meeting the developmental and educational needs of infants and preschool children with disabilities and at the same time providing mainstreaming were a complex task. How should these children be grouped for the best intervention services? When children with and without disabilities were grouped together, what

were the effects when all of them were progressing through critical periods of development? Not only was identification of young children with disabilities more complex, but evaluation of the infant and preschool programs providing intervention services was also difficult.

PL 101-576

The Americans with Disabilities Act (ADA), passed in 1990 (Stein, 1993), and the amendments to PL 94-142 (IDEA) have had an additional impact on the education of young children with disabilities. Under the ADA, all early childhood programs must be prepared to serve children with special needs. Facilities and accommodations for young children, including outdoor play environments, must be designed, constructed, and altered appropriately to meet the needs of young children with disabilities. The PL 94-142 amendments, passed in 1991, require that the individual educational needs of young children with disabilities must be met in all early childhood programs (Deiner, 1993; McCollum & Maude, 1993; Wolery, Strain, & Bailey, 1992). These laws advance the civil rights of young children and have resulted in the inclusion of young children in preschool and school-age programs. As a result, the concept of mainstreaming is being replaced by **integration**, or **inclusion**, whereby all young children learn together with the goal that the individual needs of all children will be met (Krick, 1992; Wolery & Wilbers, 1994). The efforts of these programs and their services must be assessed and evaluated to determine whether the needs of children are being met effectively.

Individuals with Disabilities Education Improvement Act of 2004

The Congress reauthorized the Education for All Children Act of 1975 in 1997 (IDEA). The reauthorization of the 1997 law required special education students to participate in state tests, and states were to report results of those tests to the public. Many states were slow to comply with the law and there were no consequences for states that did not comply.

The No Child Left Behind Act of 2001 (NCLB) required states to test at least 95% of their students with disabilities. Subsequently, the Individuals with Disabilities Education Improvement Act of 2004 was aligned with the requirements of NCLB. Final regulations of the law were officially published in August 2006. Three important rules addressed the impact of NCLB. A provision of NCLB was that highly qualified teachers must be hired. The regulations clarified this rule for special education teachers: states could create a state standard of evaluation for special education teachers.

NCLB specified that states could still use other methods of diagnosing children with learning disabilities. The response-to-intervention process involved providing intervention services for students. Students who did not respond could be referred for special education services. This process was clarified in the regulations, which stated that states could still use other methods of diagnosing children with learning disabilities. A third provision caused some controversy. This required that students in private schools would be provided services through the public schools. School districts were required to set aside a certain percentage of their federal funds for services to private school students (*Education Week*, n.d; Samuels, 2006; U.S. Department of Education, 2006).

Current Issues and Trends in Assessment in Early Childhood Education

11

CHAPTER 1
*An Overview
of Assessment
in Early
Childhood*

The 1980s brought a new reform movement in education, accompanied by a new emphasis on testing. The effort to improve education at all levels included the use of standardized tests to provide accountability for what students are learning. Minimum competency tests, achievement tests, and screening instruments were used to ensure that students from preschool through college reached the desired educational goals and achieved the minimum standards of education that were established locally or by the state education agency. As we continue in a new century, these concerns have increased.

Trends in a New Century

In the 1990s many schools improved the learning environment and achievement for all children; nevertheless, a large percentage of schools were still low performing in 2000 and 2001. Inadequate funding, teacher shortages, teachers with inadequate training, aging schools, and poor leadership affected quality education (Wortham, 2002).

During the 2000 presidential campaign, candidate George W. Bush named quality education as one of the goals of his presidency. After his election, President Bush worked for legislation that would improve education for all children. After months of dialogue and debate, Congress passed a new education act in December 2001. The No Child Left Behind Act (NCLB), signed into law on January 8, 2002, had an impact on testing required by individual states. In addition to other provisions, all states were required to administer tests developed by the state and to set and monitor adequate yearly progress (Moscosco, 2001; Wortham, 2002).

President Bush was also committed to strengthening early childhood programs. In 2002, several projects were conducted to support early childhood programs. Under the Sunshine Schools program, the U.S. Department of Education focused on what is working in early childhood education and gave attention to highly effective state, district, city, county, and campus programs (Grissom, personal communication, April 4, 2002).

Another Bush initiative, Good Start, Grow Smart, was intended to strengthen Head Start and improve the quality of experiences for children. The initiative provided the following:

- Training for nearly 50,000 Head Start teachers on the best techniques
- Assurance that preschool programs are more closely coordinated with K–12 educational programs
- A research effort to identify effective early literacy programs and practices (Grissom, personal communication, April 4, 2002).

In July 2001, the White House hosted the White House Summit on Early Childhood Cognitive Development. The Early Childhood–Head Start Task Force formed following the summit published a new guide, *Teaching Our Youngest* (Grissom, personal communication, April 4, 2002).

The early childhood education projects initiated by the Bush administration to improve education stressed the importance of improving early childhood programs;

The No Child Left Behind Act of 2001

NCLB requires states to do the following (U.S. Department of Education, 2001):

- Provide public school choice and supplemental services for students in failing schools as early as fall 2002.
- Integrate scientifically based reading research into comprehensive instruction for young children.
- Set and monitor adequate yearly progress, based on baseline 2001–2002 data.
- Issue annual report cards on school performance and statewide test results by 2002–2003.
- Implement annual, standards-based assessments in reading and math for grades 3 to 8 by 2005–2006.
- Assure that all classes are taught by a qualified teacher by 2005–2006.

U.S. Department of Education (2001). Retrieved February 14, 2007, from <http://www.ed.gov/aclb/overview/intro/factsheet/html>.

nevertheless, there is no doubt that mandates for increased standards-based testing will continue in the future in spite of concerns of their relevancy, especially for young children. Fortunately, child-outcome standards have also been developed by professional organizations in addition to state education agencies. The National Council for the Social Studies issued *Curriculum Standards for the Social Studies* (National Council for the Social Studies, 1994). Improved Head Start Performance Standards published in 1996 included children from birth to age 5 (Early Head Start, 2000). These standards and others provide guidelines for early childhood educators as they strive to improve programs and experiences for young children. By 2005, standards that included early childhood were available in many states. Some were in response to NCLB, but others were part of the emerging efforts to establish state and national standards for development and learning (Seefeldt, 2005).

Individual states are continuing to develop, implement, and review early learning guidelines as the set standards for preschool curriculum. All states except for Hawaii were engaged in or had completed the process in 2009 (National Child Care Information and Technical Assistance Center [NCCIC], 2009).

The Accountability Era

The major issue in education today is the idea of accountability. Even before the rules and regulations surrounding the legislation for No Child Left Behind (NCLB) were issued, there were growing concerns about accountability. The interest in developing more responsibility for student results evolved from a perception that

states had been evaluating school systems on the basis of available resources rather than student performance. NCLB addressed student performance, public reporting of achievement results, consequences for poor student performance, and continuous improvement (Edweek, 2004). Individual states were also responding to the need for accountability by moving from a focus on curriculum offerings and funding levels to standards-based accountability. States now have set standards, developed assessment systems, and assigned responsibilities for meeting the goals and designating rewards and sanctions to achievement levels. If states want to continue getting benefits under NCLB, they have to follow the new policies for accountability (National Council of State Legislatures, 2009).

Emerging Issues With NCLB

The requirements of NCLB were to be implemented by 2006. In the summer of 2006 it was evident that there were difficulties in complying with the law.

An early issue was the requirement that schools report test scores by racial subgroup. Nearly two dozen states had been granted waivers in reporting by subgroups. Other schools avoided the problem by determining that numbers of students in racial subgroups were too small to be statistically significant. Their scores were not included (Rebora, 2006).

The law also provided that states would implement standards-based assessments in reading and math by 2006. Ten states were notified in 2006 that a portion of state administrative funds would be withheld for failing to comply fully with NCLB. Twenty-five states might also lose a portion of their aid if they didn't comply fully with NCLB and comply with the testing requirement by the end of the school year. The monetary penalties caught many states by surprise. In addition, states had difficulty providing the extensive documentation required to demonstrate that the tests met that state's academic standards (Olson, 2006). Further, states had to



Assessments can be conducted while young children engage in independent work.

demonstrate how they were including students with disabilities and English language learners (ELLs) in their testing system. This included developing alternative assessments when needed. When combined with concerns about testing young children in the early childhood years, NCLB had an impact on all populations of students, including those in the preschool years.

The reauthorization of NCLB was due in 2007. Congress had already blocked action on the reauthorization until after the 2008 election. The Obama administration indicated in 2009 that the rewriting of the law would focus on teacher quality, academic standards, and more attention given to help failing schools and students. The Commission on No Child Left Behind (2009) urged Secretary of Education Arne Duncan to retain some core elements of NCLB. Regardless of the direction of continuing reform in education, the federal government would continue to expand its influence on accountability and would also encourage the movement from individual state standards to national standards (Dillon, 2009; *The New York Times*, 2009).

Concerns About Testing Young Children in Early Childhood Settings

The increased use of testing at all levels has been an issue in American education, but the testing of young children is of particular concern. Standardized tests and other assessment measures are now being used in preschool, kindergarten, and primary grades to determine whether children will be admitted to preschool programs, promoted to the next grade, or retained. During the late 1980s and early 1990s, tests were used to determine whether students should be promoted from kindergarten to first grade or placed in a “transitional” first grade. Although this practice is now less popular, it persists in some school districts and states (Smith, 1999). In 2000, the National Association of Early Childhood Specialists in State Departments of Education (NAECS/SDE) was concerned about the continuing trend to deny children’s entry to kindergarten and first grade. They issued a position statement, “Still! Unacceptable Trends in Kindergarten Entry and Placement” (National Association of Early Childhood Specialists in State Departments of Education [NAECS/SDE], 2000). This continuing effort to advocate appropriate assessment of very young children was endorsed by the Governing Board of the National Association for the Education of Young Children (NAEYC, 2001).

By 2006, states used a wide range of types of assessments with young children entering public school. Screening tests were in use in many states for hearing and vision as well as developmental assessments and readiness tests. Many states conducted screening to identify children at risk for failing to succeed in school and/or developmental disorders or disabilities. Some states met the criteria for developmentally appropriate assessments, while others did not. For example, California required observation and portfolio materials in preschool assessments. On the other hand, Georgia students were tested for first-grade readiness at the end of the kindergarten year to determine grade placement (Education Commission of the States, 2006). More information on these topics will be provided in later chapters.

The announcement by President Bush in 2003 that all Head Start students would be given a national standardized test assessment raised new concerns. At issue were validity and reliability of tests for preschool children (Nagle, 2000) and whether such “high-stakes” testing should be used to evaluate the quality of Head Start programs (Shepard et al., 1998). Policy makers had to address these and other concerns about appropriate assessment of young children in their decisions about how to evaluate preschool programs that receive federal funding (McMaken, 2003).

In February 2003, a large group of early childhood experts wrote to their congressional representatives to express their concerns about the impending test. They made the following points:

1. The test is too narrow.
2. The test may reduce the comprehensive services that ensure the success of Head Start.
3. The test is shifting resources away from other needs within Head Start.
4. Testing should be used to strengthen teaching practices, not evaluate a program, and should in no way be linked to program funding (Fair Test, 2003; NAEYC, 2004).

In September 2003, the new test, the *National Reporting System* (NRS) (U.S. Department of Health and Human Services [HHS] Head Start Bureau, 2003), was administered by the Head Start Bureau in the Department of Health and Human Services (HHS) Administration for Children and Families to more than 400,000 children ages 4 and 5, and continues to be administered each year. In 2005, when Head Start funding was being considered, the Government Accountability Office (GAO) issued a report on the NRS. The report said that the NRS had not shown that it provided reliable information on children’s progress during the Head Start program year, especially for Spanish-speaking children. Moreover, the NRS had not shown that its results were valid measures of the learning that took place in the program. In its recommendations, the GAO required that the Head Start Bureau establish validity and reliability for the NRS. As a result the NRS was not to be used for accountability purposes related to program funding (Crawford, 2005; Government Accountability Office [GAO], 2005). Because the Bush administration reportedly intended to use the NRS to establish accountability requirements similar to NCLB, this GAO finding essentially halted the use of the test for that purpose.

Concerns About Testing Young Children With Cultural and Language Differences

A concurrent concern related to current trends and practices in the assessment of young children is the question of how appropriate our tests and assessment strategies are in terms of the diversity of young children attending early childhood programs. Socioeconomic groups are changing dramatically and rapidly in our society, with an expansion of the poorer class and a corresponding shrinking of the middle class (Raymond & McIntosh, 1992). At the same time, an increase in minority citizens has occurred as the result of the continuing influx of people from other

countries, especially Southeast Asia and Central and South America. Moreover, Hispanic families are no longer concentrated in the Southwest; their growth in many parts of the country has caused new communities to have unprecedented high percentages of Hispanic children. Seventy-nine percent of young ELLs in public schools speak Spanish. In addition, approximately 460 languages are represented in schools and programs in the United States, including Spanish, Chinese, Arabic, Armenian, and Hmong (Biggar, 2005; Lopez, Salas, & Flores, 2005). Assessment of the developmental progress of children from these groups is particularly important if their learning needs are to be identified and addressed.

Evidence shows that standardized test scores have had a high correlation to parents' occupations, level of education, the location of the student's elementary school, and the family's income bracket. Moreover, students from limited English backgrounds tend to score lower on reading and language fluency tests in English. They typically perform better on computational portions of mathematics tests (Wesson, 2001). The fairness of existing tests for children who are school disadvantaged and linguistically and culturally diverse indicates the need for alternative assessment strategies for young children (Biggar, 2005; Goodwin & Goodwin, 1993, 1997). A major issue in the 21st century is appropriate measurement and evaluation strategies that will enhance, rather than diminish, the potential for achievement.

The history of assessment of minorities who are bilingual students or learning English as a second language is one of potential bias. Children have been and continue to be tested in their nondominant language (English) or with instruments that were validated on an Anglo, middle-class sample of children. As a result, many Hispanic preschool children were and are still regularly diagnosed as developmentally delayed and placed in special education (Lopez et al., 2005). The issue of appropriate assessment of these children was addressed by court cases such as *Diana v. California State Board of Education* (1968) and *Lau v. Nichols* (1974). More recently, NCLB and the Head Start NRS have addressed the issue of testing ELLs (Crawford, 2005; David, 2005; GAO, 2005).

The overidentification of minority students for special education is often related to language and cultural differences. Some of the issues addressed in the rising numbers of minority children being referred to special education were traced in one study to inconsistent methods of determining home language and English proficiency, confusion as to the purpose of language screening instruments, and a need for more training for teachers in meeting the needs of culturally and linguistically diverse children and families (Abebe & Hailemariam, 2008; Hardin, Roach-Scott, & Peisner-Feinberg, 2007).

Increasing concerns about overidentification of minority children is addressed in two significant books. *Why Are So Many Minority Students in Special Education? Understanding Race and Disability in Schools* (Harry & Klingner, 2005) is one effort to explain the problem. The authors address the issue of the disproportionate representation of minorities in special education. *Racial Inequity in Education* (Loren & Orfield, 2002) addresses many factors that include language, high-stakes testing, inappropriate and inadequate special education for minority children, and the role of the federal government.

Another concern about testing children with cultural and language differences is the process of screening preschool children who fit into this category. A problem of correctly screening young children who are learning English may lead to the

underidentification of children who have special needs or overidentification of special needs because English language delays are misdiagnosed as a disability (NAEYC, 2005a). Recommendations were made for appropriate screening and assessment and program accountability for correctly serving young children in English.

The impact of NCLB on testing ELLs has resulted in the development of new English language proficiency tests based on new standards adopted by each state. More importantly, the tests measure the reading, writing, speaking, and listening skills of ELLs (Zehr, 2006). In summer 2006, five states had failed to meet the Department of Education's deadline to have tests in place. While some states designed their own tests, other states adopted tests designed by consortia or testing corporations. Nevertheless, because test development and implementation were still in the beginning stages, little was known about the validity and reliability of the tests and whether the tests met the requirements of the law. The New York example reveals the complexity of the assessment of ELLs. The New York State test was designed to measure language acquisition, while the tests meeting NCLB measured English language skills. This was true for bilingual and ELL programs throughout the United States prior to NCLB. It would take many years to develop and validate tests that would resolve how to assess the language skills of limited-English speakers that were comparable with tests for English-speaking students.

Assessment of young children who are from families that are culturally and linguistically diverse must include many dimensions of diversity. It is not useful to proceed with assessment that is culturally fair for Hispanic or Asian populations generally. The many variations within communities and cultures must be considered, among them the educational background of the parents and the culture of the immediate community of the family. Congruence between the individual cultural perceptions of the assessors and the children being assessed, even when both are from the same culture or language population, must also be considered (Barrera, 1996). Many types of information, including the child's background and the use of assessments, must be combined to determine a picture of the child that reflects individual, group, and family cultural characteristics (Lopez et al., 2005).

Concerns About Testing Young Children With Disabilities

The use of testing for infants and young children with disabilities cannot be avoided. Indeed, Meisels, Steele, and Quinn-Leering (1993) reflected that not all tests used are bad. Nevertheless, Greenspan, Meisels, and others (1996) believe that assessments used with infants and young children have been borrowed from assessment methodology used with older children and do not represent meaningful information about their developmental achievements and capacities. Misleading test scores are being used for decisions about services, educational placements, and intervention programs. These developmental psychologists propose that assessment should be based on current understanding of development and use structured tests as one part of an integrated approach that includes observing the child's interactions with trusted caregivers. Assessment should be based on multiple sources of information that reflect the child's capacities and competencies and better indicate what learning environments will best provide intervention services for the child's optimal development.

Play-based assessment is one major source of information among the multiple sources recommended. Play assessment is nonthreatening and can be done unobtrusively. Moreover, during play, children can demonstrate skills and abilities that might not be apparent in other forms of assessment. Children's ability to initiate and carry out play schemes and use play materials can add significant information (Fewell & Rich, 1987; Segal & Webber, 1996). In transdisciplinary play-based assessment, a team that includes parents observes a child at play. Each member of the team observes an area of development. During the assessment the child's developmental level, learning styles, patterns of interaction, and other behaviors are observed (Linder, 1993).

NCLB has had an impact on curriculum and assessment of children with disabilities. While identification of children can begin very early in life, the needs of the children as they enter public education are not usually identified until first grade. However, during the last 10 years, the nature and objectives of kindergarten have changed because of advances in knowledge about what young children are capable of learning and the advent of the standards-based accountability movement. Kindergarteners are taught and tested on the mastery of academic standards. This change in expectations has affected the kindergarten year for children at risk for learning disabilities. The kindergarten year formerly was used to work with at-risk children and refer them for testing at the end of the year. When they reached first grade they would be referred for identification and possible special education services. Children with disabilities or who are at risk for learning problems now need identification and services earlier than first grade. Identification of disabilities and referral for services should now be considered for the kindergarten year, even if some disabilities are difficult to identify in early childhood (Litty & Hatch, 2006).

NCLB also added accountability measures to IDEA, as described earlier in the chapter. School districts must test at least 95% of students with disabilities and incorporate their test scores into school ratings. There has been strong public reaction to the inclusion of special education students in state testing and reporting. Some policy makers see this provision as an important step in every child receiving a high-quality education. Critics worry that the law is not flexible enough to meet individual needs of students with disabilities. Many teachers felt that special education students should not be expected to meet the same set of academic content standards as regular education students. These issues were yet to be resolved when the final regulations were published in August 2006 for the Individuals with Disabilities Education Improvement Act of 2004 (*Education Week*, n.d.; U.S. Department of Education, 2006).

Since 2006, work has continued to address the issue of identifying and serving students with learning disabilities. The focus of this effort has been to find more flexible and research-based strategies for both identifying students who need intervention services and better serving students with quality instruction and evaluation (Division for Early Childhood of the Council for Exceptional Children, 2007). Two models for a more inclusive instructional process for all students are Response to Intervention (RTI) and Universal Design for Learning (UDL).

Response to Intervention addresses all student needs whether or not they have been identified as learning disabled. RTI is implemented through a three-tiered process of responding to the needs of all children (Burns & Coolong-Chaffin, 2006; Millard, 2004). All students begin at the first tier. Students who need more targeted

education are served in the second tier. Students who need intensive intervention are served in the third tier. This tier can include special education services.

The RTI model seeks to match students with the most effective instruction. The core features of RTI are high-quality classroom instruction, research-based instruction, classroom performance, universal screening, continuous progress monitoring during interventions, and fidelity measures (Millard, 2004).

Universal Design for Learning (UDL) also seeks to include all kinds of students, including students with learning disabilities, English language barriers, emotional or behavior problems, lack of interest or engagement, or sensory and physical disabilities. UDL is based on the need for multiple approaches to instruction that meet the needs of diverse students (Center for Applied Special Technology [CAST], 2009). It applies recent research on neuroscience and uses technology to make learning more effective for all students. The curriculum includes customized teaching that includes multiple means of representation, multiple means of action and expression, and multiple means of engagement (CAST, 2009).

Authentic and Performance Assessment

Assessment is in a period of transition. Teachers of young children are moving from more traditional strategies of assessing for knowledge and facts to assessing the students' ability to reason and solve problems. Despite the demands for accountability for addressing early childhood standards, assessments provide a variety of methods for children to demonstrate what they understand and can do.

A broader view of assessment has incorporated a multidimensional approach to measurement, as described earlier in the sections on concerns for assessment of children from diverse populations and children with disabilities. It is now felt that too much attention has been given to the use of standardized tests, rather than a multidimensional approach that uses many sources of information. The more inclusive practice of assessment, which includes work samples, observation results, and teaching report forms, is called **alternative assessment**. These alternatives to standardized tests measure how students can apply the knowledge they have learned (Blum & Arter, 1996; Maeroff, 1991). Within this evolution in the purposes for assessment and interpretation of assessments is the move to authentic and performance assessments. **Authentic assessments** must have some connection to the real world; that is, they must have a meaningful context. They are contextual in that they emerge from the child's accomplishments. **Performance assessments** permit the child to demonstrate what is understood through the performance of a task or activity (Wortham, 1998).

Performance assessment as applied through the use of portfolios provides a multifaceted view of what the young child can understand and use. Performance assessment is used because teachers in early childhood programs seek information about the child's development and accomplishments in all domains. Performance assessment combined with other assessments provides a longitudinal record of change in development, rather than an assessment of a limited range of skills at a particular time. It is appropriately used with infants, young children, school-age children, children from diverse populations, and children with disabilities (Barrera, 1996; Meisels, 1996; Wortham, 1998).

Documentation is another form of performance assessment. First developed in Reggio Emilia schools in Italy and now widely used in the United States, documentation is a process of collecting and displaying children's work on projects (Wurm, 2005). More about documentation will be discussed in chapter 8.

This broader view of assessment in early childhood programs is echoed by the organizations that endorsed and supported the *Guidelines for Appropriate Curriculum Content and Assessment in Programs Serving Children Ages 3 Through 8*, a position statement of the NAEYC and the NAECs/SDE adopted in 1990 and renewed in 2000 and 2001 (NAEYC, 1992; NAECs/SDE, 2000). These guidelines proposed that the purpose of assessment is to benefit individual children and to improve early childhood programs. Appropriate assessment should help enhance curriculum choices, help teachers collaborate with parents, and help ensure that the needs of children are addressed appropriately. Rather than being narrowly defined as testing, assessment should link curriculum and instruction with program objectives for young children (Hills, 1992). Authentic and performance assessments provide dynamic assessment approaches that benefit the child, parents, caregivers, and teachers.

Standards for Beginning Teachers

The era of accountability includes expectations for the appropriate preparation of teachers. Just as states set standards for student curriculum and assessment for diverse children, there are standards for preparing and assessing whether beginning teachers are qualified to teach young children.

The Interstate New Teacher Assessment and Support Consortium (INTASC) includes state education agencies and national education organizations. The consortium believes that each state's education system should have a teacher licensing policy that requires teachers to know and be able to effectively help all students achieve the state standards for students (Council of State School Officers, 2007, 2009).

The Mission of INTASC

The mission of INTASC is to provide a forum for its member states to learn and collaborate in the development of

- Compatible educational policy on teaching among the states.
- New accountability requirements for teacher preparation programs.
- New techniques to assess the performance of teachers for licensing and evaluation.
- New programs to enhance the professional development of teachers (Council of Chief State School Officers, 2007, p. 1).

The licensing standards for early childhood teachers has been addressed by three organizations: the Association of Teacher Education (ATE), the National Association for the Education of Young Children (NAEYC), and the Association for Childhood Education International (ACEI). A position statement on early childhood teachers was issued by ATE and NAEYC in 1991 (ATE & NAEYC, 1991). The position statement also calls for state early childhood organizations and agencies to develop policies leading to certification that is distinct from policies related to elementary and secondary certification. In addition, policies for early childhood teachers should be congruent across the 50 states.

The *Position Paper on the Preparation of Early Childhood Education Teachers* was issued by ACEI in 1998 (Association for Childhood Education International [ACEI], 1998). It calls for early childhood specialization to be developed within broader policies for teacher preparation. Early childhood teachers should have a broad and liberal education. Experiences should also include foundations of early childhood education, child development, the teaching and learning process, and provisions for professional laboratory experiences.

NAEYC also developed a position statement on ethical conduct (NAEYC, 2005). Standards of ethical behavior by early childhood care and education teachers are based on a commitment to

- Appreciate childhood as a unique and valuable stage of the human life cycle.
- Base our work on knowledge of how children develop and learn.
- Appreciate and support the bond between child and family.
- Recognize that children are best understood and supported in the context of family, culture, community, and society.
- Respect the dignity, worth, and uniqueness of each individual (child, family member, and colleague).
- Respect diversity in children, families, and colleagues.
- Recognize that children and adults achieve their full potential in the context of relationships that are based on trust and respect (NAEYC, 2005b, p. 1).

Summary

The measurement and assessment of children begins very early in the life span. Newborns are tested for their neonatal status, and infant tests designed to assess development begin the trend for testing and assessment in the early childhood years. Assessments in the early childhood years have many purposes; some are beneficial for young children, and others are detrimental.

The advent of measures to assess and evaluate young children's development and learning occurred at the beginning of the 20th century. As the decades passed, significant trends in the study of young children and services and programs implemented for young children have driven the need to develop standardized tests and other measures to evaluate children's progress and program effectiveness.

Many issues surround the testing of young children. Some educators question the validity and reliability of standardized tests used with young children, as well as the purposes for administering tests to children who are culturally and linguistically diverse. At the same time, the use of individual testing and evaluation to identify children with disabilities and provide services for them continues to serve a valuable purpose.

REVIEW QUESTIONS

1. Why are very young children measured in infancy and in the preschool years? Give examples.
2. Explain developmental deficits. How are developmental deficits identified and treated?
3. Why is research conducted on the development of very young children? How can such research be used?
4. How were Pestalozzi and Rousseau pivotal in the origins of understanding and measuring young children?
5. Why has the child study movement been the major resource for understanding child development?
6. How does the history of standardized testing include testing with infants and young children? What kinds of standardized tests are beneficial for children under age 6?
7. Why were standardized tests developed for Head Start? How were they used?
8. Why were standardized tests developed as a result of legislation for young children with disabilities? How are they used?
9. Why is it difficult to develop assessments for children who are culturally and linguistically different? What factors must be addressed in their assessment?
10. What are some of the weaknesses in assessments of young children with disabilities? How can these difficulties be overcome?
11. How is authentic assessment different from assessment using standardized tests?

SUGGESTED ACTIVITIES

1. Review a recent journal article on a topic related to current issues in the testing and assessment of young children. The article should have been published within the past 5 years. Describe the major points in the article and your response. Be prepared to share in small groups.
2. What are the policies followed in your state regarding the use of standardized tests? What tests are administered in the primary grades? How are they chosen? How are the results used?
3. How does the school district in your community screen preschool children for possible disabilities? What types of assessments are used? If children need further testing to identify specific needs, what process is used? Who conducts the tests with the child?

KEY TERMS

alternative assessment
authentic assessment
documentation
inclusion

integration
mainstreaming
performance assessment