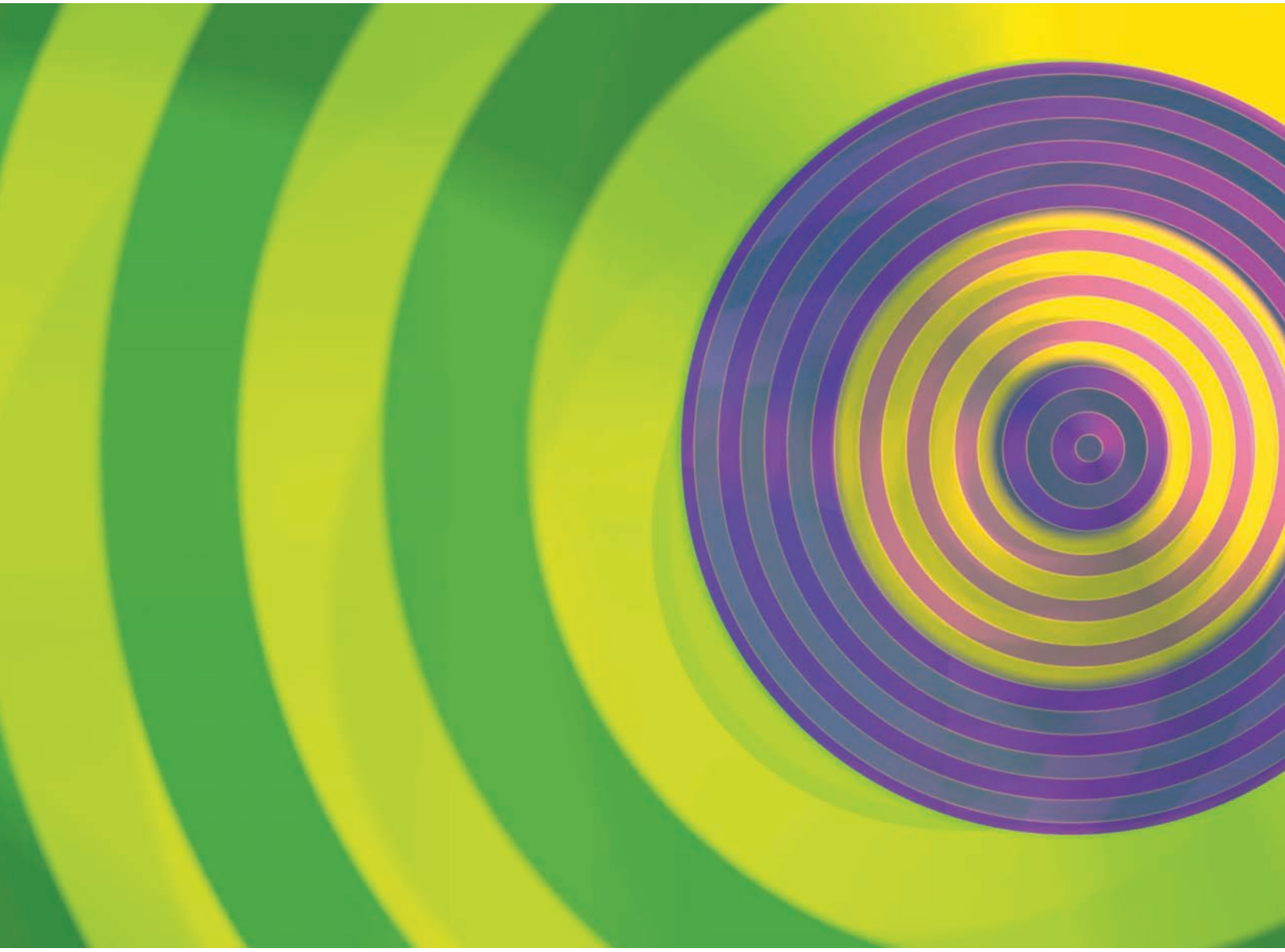
Pearson New International Edition



Assessment in Early Childhood Education
Sue C. Wortham
Sixth Edition

ALWAYS LEARNING™

PEARSON®

Pearson New International Edition

Assessment in Early Childhood Education
Sue C. Wortham
Sixth Edition

PEARSON®

Tests are administered to young children to acquire beneficial information about them. In chapter 3, we discussed how standardized tests are planned, designed, and standardized.

In this chapter, we discuss in more detail how to use information from children's test scores. In the process of standardizing a test, developers establish the norms that make test score interpretation useful. We not only take a more detailed look at norm-referenced tests, but also study how another type of standardized test, the criterion-referenced test, is used to meet the learning needs of young children. Group test scores can be used to analyze and improve curriculum and instruction at various levels within a school district; individual test scores can be used by the classroom teacher to organize appropriate learning experiences for individual students or the class as a whole.

We also discuss how individual and group test results are used to report student progress and program effectiveness. Test results are important to teachers, school district administrators, parents, and school boards. Results are reported to each in a context that provides meaningful interpretation of the test. Finally, we consider the disadvantages and advantages of using norm- and criterion-referenced tests with young children.

Uses of Norm-Referenced and Criterion-Referenced Tests

Distinctions Between Norm-Referenced and Criterion-Referenced Tests

Norm-referenced and criterion-referenced tests are both standardized instruments. Some standardized tests are designed for norm-referenced results and others for criterion-referenced results. The current trend is to design tests that are both norm and criterion referenced. The two types of tests have different purposes, and test items are used differently when measuring student learning or achievement. **Norm-referenced tests** provide information on how the performance of an individual compares with that of others. The individual's standing is compared with that of a known group. The person's percentile rank is obtained to determine the relative standing in a norm group by recording what percentage of the group obtained the same score or a lower score.

In contrast, **criterion-referenced tests** provide information on how the individual performed on some standard or objective. These test results allow users to interpret what an individual can do without considering the performance of others. Criterion-referenced tests are designed to measure the results of instruction; they determine the individual's performance on specific behavioral or instructional objectives (Wilson, 1980; Zucker, 2003). Linn and Miller (2005) describe the difference between the two types of tests as the ends of a continuum: "The criterion-referenced test emphasizes description of performance and the norm-referenced test emphasizes discrimination among individuals" (p. 44).

Regardless of whether tests are norm or criterion referenced, the process of their design and development is as described in chapter 3. They are constructed and standardized through all the steps that will result in validity and reliability. It is also possible that norm- and criterion-referenced tests have not been standardized; however, criterion-referenced tests are more often nonstandardized (Goodwin & Goodwin, 1993). It is equally important that criterion-referenced tests have validity and reliability if they are to be used to make decisions about young children.

Norm- and criterion-referenced tests have characteristics in common. Linn and Miller (2005) describe these as follows:

1. Both require a relevant and representative sample of test items.
2. Both require specification of the achievement domain to be measured.
3. Both use the same type of test items.
4. Both use the same rules for item writing (except for item difficulty).
5. Both are judged by the same qualities of goodness (validity and reliability).
6. Both are useful in educational measurement. (p. 14)

Both tests measure what students have learned; nevertheless, the objectives for measurement are different. The norm-referenced test is broad in content. Many aspects of the content are measured. Because the test is concerned with overall achievement, only a small sample of behaviors for each objective can be assessed. The criterion-referenced test focuses on mastery of objectives. Each objective has many test questions to determine whether the objective has been mastered (Zucker, 2003).

An achievement test in mathematics provides a good example. The norm-referenced test for the first grade may have items on addition, subtraction, sets, and all other areas included in the mathematics curriculum. Test items are written to sample the student's overall performance in first-grade mathematics. The student's total raw score is then transformed to compare overall achievement with the test norms. On the criterion-referenced test, student performance on individual curriculum objectives is important. Test items are written to measure whether the child has mastered a particular learning objective in subtraction, addition, or other components of the mathematics curriculum (Goodwin & Goodwin, 1982).

Another difference between norm- and criterion-referenced tests also relates to differences in test items. In a norm-referenced instrument, test items must cover a wide range of difficulty. Because the test is intended to discriminate between the performance of students and groups of students, the difficulty of test items ranges above the grade level for which the test is intended. Test items designed primarily for criterion-referenced purposes are written specifically for learning tasks. Easy items are not omitted, and the intent is to evaluate how well the student has learned the objectives for one grade level (Wilson, 1980).

New standardized tests have been developed with dual referencing; that is, they are designed for both norm- and criterion-referenced assessment. Although it is difficult to develop a single test that works equally well for both types of measurements, obtaining both kinds of performance results is helpful to educators. Compromises in test construction are offset by the more effective use of the test (Linn & Miller, 2005). Some criterion-referenced tests have not been standardized. This does not imply that they are not well designed and useful, but readers should be aware of this condition.

Uses of Norm-Referenced Tests With Preschool Children

Norm-referenced test scores are used to measure individual achievement within a designated group. Norms are not standards to be reached; they are numerical descriptions of the test performance of a group of students. Norms can be established at a national or local level. Norm-referenced tests are commonly used to measure school achievement, intelligence, aptitude, and personality traits. Formal tests are administered at the preschool level to identify children who need or can benefit from special instruction, as well as to determine the success of an early childhood program.

Measures of intelligence such as the *Wechsler Preschool and Primary Scale of Intelligence* (Wechsler, 2002) are norm-referenced instruments that allow test examiners to differentiate the knowledge skills of preschool students. As discussed in chapter 2, intelligence tests are described as diagnostic because they include comprehensive examination of children who might be mentally or physically delayed or who are at risk for learning disabilities. Other tests in this category include the *Kaufman Assessment Battery for Children (K-ABC-II)* (Kaufman & Kaufman, 2005) and *McCarthy's Scales of Children's Abilities* (McCarthy, 1983). In addition to identifying children with disabilities, intelligence tests can be used to identify children who are gifted.

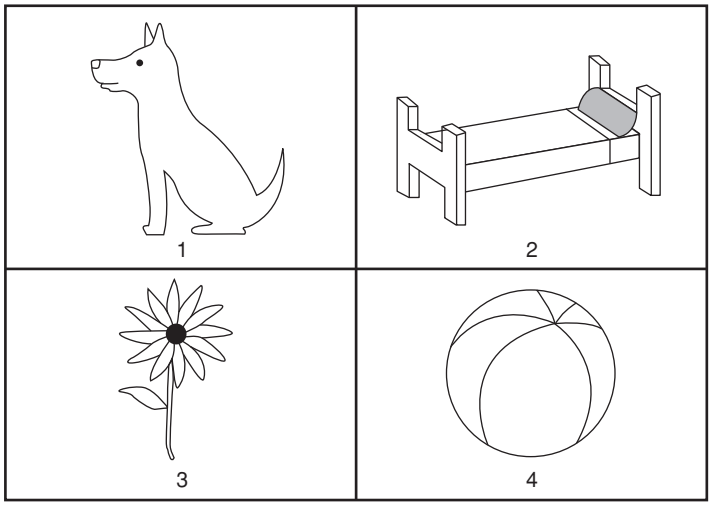
The National Reporting System (U.S. Department of Health and Human Services, 2003) discussed in earlier chapters is a controversial norm-referenced test implemented with the purpose of establishing accountability in Head Start programs. Efforts to use test results to determine whether programs would be refunded in 2005 failed. (See chapter 1.) The test is administered orally to individual children. Figure 4-1 shows sample questions on vocabulary and early math.

Norm-referenced tests are used with preschool children to measure their present level of knowledge, skills, or performance. In federally funded programs such as Head Start, a norm-referenced measure may be used to evaluate the learning acquired by the children as a result of the program. The *Peabody Picture Vocabulary Test* (Dunn & Dunn, 1997) provides a measure for language development. The *Boehm Test of Basic Concepts* (Boehm, 2000) and the *Learning Accomplishments Profile—Revised* (Sanford & Zelman, 1995) assess the child's abilities and skills, including the acquisition of concepts.

Uses of Norm-Referenced Tests With School-Age Children

After children enter primary school, achievement tests are the most frequently administered norm-referenced tests. Locally developed achievement tests, as well as state and national tests, can be given in order to measure and analyze individual and group performance resulting from the educational program. Children experiencing difficulties in school are evaluated with screening and diagnostic tests, but all

VOCABULARY
I have said. Let's try one. Put your finger on "ball."
IF THE CHILD RESPONDS CORRECTLY WITHOUT HELP BY POINTING TO THE BALL IN QUADRANT 4 SAY:
Good! Let's try another one. Put your finger on "dog."
(SEE GRAPHIC BELOW.)



EARLY MATH addresses child outcomes in numbers and operations. The math skills assessed include recognizing one-digit numerals and basic geometric shapes, solving word problems involving counting or simple addition or subtraction, and interpreting simple measurements. The estimated administration time is 3 minutes.

How many trees are on this page?
CORRECT: TWO (NONVERBAL RESPONSES ARE ACCEPTABLE)
(SEE GRAPHIC BELOW)

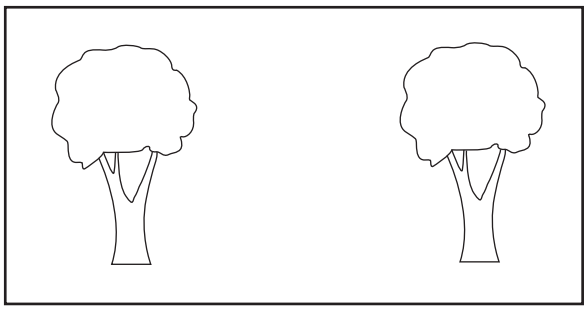


FIGURE 4-1 Sample questions for the Head Start National Reporting System

Source: U.S. Department of Health and Human Resources Head Start Bureau. (2003). *National Reporting System.* Washington, DC: Author.

students take achievement tests as early as kindergarten, more frequently beginning in first grade.

Norm-referenced test results are used for more general comparisons of group test results. One such use is to assess achievement level in subject areas. The achievement of a single class in a school, all classes of a certain grade level in the school, all schools at a grade level in a school district, and all schools within a state with that grade level can be studied to determine general progress in one or more subject areas. The results of batteries of tests can be analyzed for trends in achievement.

In a similar type of analysis, components of an instructional program can be studied by using group test scores. If a new instructional program is to be tried or if an existing method is to be evaluated to help in deciding whether changes are needed, an achievement test can be used to investigate the effectiveness of the program. Particular areas of weakness and strength can be pinpointed, and decisions and plans can be made to improve weak components in the curriculum.

Uses of Criterion-Referenced Tests With Preschool Children

Criterion-referenced test scores are used to describe individual performance on specific objectives. Criterion-referenced measures de-emphasize distinctions among individual performances; rather, they indicate whether the individual has mastered the objectives that were tested. Criterion-referenced tests are used for developmental screening, **diagnostic evaluation,** and instructional planning.

In the preschool years, developmental and diagnostic assessments are the criterion-referenced tests used most frequently. Although **developmental screening** is used primarily to identify children who might profit from early education intervention or from special services before kindergarten or first grade, it is also used as a checkpoint for children who are developing normally. The *Dynamic Indicators of Basic Early Literacy Skills (DIBELS)* (Good & Kaminski, 2002) is an example of a screening for literacy skills. One of the subtests, the *DIBELS Individual Sound Fluency (ISF)* is a measure of phonological awareness. Figure 4-2 shows a sample of the test in which the student is asked to identify/produce the correct beginning sound of a word. This subtest is administered to preschool and kindergarten children.

As introduced in chapter 3, various screening tests have been developed as a result of Public Law 94–142, the Individuals with Disabilities Education Act, which required children with disabilities to be placed in the "least restrictive environment" possible. As described by Meisels (1994), "Early childhood developmental screening is a brief assessment procedure designed to identify children who, because of the risk of a possible learning problem or handicapping condition, should proceed to a more intensive level of diagnostic assessment" (p. 1). Thus, developmental surveys assess affective, cognitive, and psychomotor characteristics to determine whether further testing evaluation is needed to identify disabilities and strategies for remediation.

Various screening tests have been developed for the preschool child. The *Denver II* (Frankenburg, Dodds, Archer, Shapiro, & Bresnick, (1990) is commonly

5. Correct Initial Consonant Sound: If the word starts with an initial consonant sound, the child can respond with the first sound or initial sounds. For example, if the word is "clock" a correct initial sound would be /c/ or /cl/ or /klo/ but not /l/ or "clock."

PROMPT:	***STUDENT SAYS:***	***SCORE:***
What sound does "clock" begin with?	/k/	0 (1)
What sound does "clock" begin with?	/kl/	0 (1)
What sound does "clock" begin with?	/klo/	0 (1)
What sound does "clock" begin with?	/l/	(0) 1
What sound does "clock" begin with?	"clock"	(0) 1

6. Correct Initial Vowel Sound: If the word starts with an initial vowel sound, the child can respond with the initial vowel sound or initial sounds. For example, if the word is "elephant" a correct initial sound would be /e/ or /el/ or /ele/, but not the name of the letter /ea/.

PROMPT:	***STUDENT SAYS:***	***SCORE:***
What sound does "elephant" begin with?	/e/	0 (1)
What sound does "elephant" begin with?	/el/	0 (1)
What sound does "elephant" begin with?	/ea/	(0) 1
What sound does "elephant" begin with?	/ele/	0 (1)

7. Schwa sound (/u/) added to a consonant is not counted as an error. Some phonemes cannot be pronounced correctly in isolation without a vowel, and some early learning of sounds includes the schwa.

PROMPT:	***STUDENT SAYS:***	***SCORE:***
What sound does "clock" begin with?	/ku/	0 (1)
What sound does "clock" begin with?	/klu/	0 (1)

8. Articulation Difficulty: The student is not penalized for imperfect pronunciation due to dialect, articulation, or second language interference. For example, the student responds /th/ when asked for the first sound in "sink". If the student consistently say /th/ for /s/, as in "thircle" for "circle," he or she should be given credit for a correct initial sound. This is a professional judgment and should be based on the student's responses and any prior knowledge of his/her speech patterns.

PROMPT:	***STUDENT SAYS:***	***SCORE:***
What sound does "sink" begin with?	/th/	0 (1)

FIGURE 4-2 Examples of questions from the DIBELS test

Source: Good, R. H., & Kaminski, R. A. (Eds.). (2002). *Dynamic Indicators of basic Early Literacy Skills: Administration and scoring guide* (6th ed.). Eugene, OR: Institute for the Development of Educational Achievement. Reprinted by permission of Dynamic Measurement Group.

used by pediatricians and other medical professionals. The *Early Screening Inventory—Revised* (Meisels, Marsden, Wiske, & Henderson, 2008) and *McCarthy's Scales of Children's Abilities* (McCarthy, 1978, 1983) are also used for screening purposes. Figure 4-3 shows some of the criterion-referenced screening items on the *Early Screening Inventory—Revised.*