Norm-Referenced and Criterion-Referenced Tests
Use in Pediatrics and Application to Task Analysis of Motor Skill

PATRICIA C. MONTGOMERY
and BARBARA H. CONNOLLY

The purposes of this article are 1) to compare the similarities and differences between norm-referenced and criterion-referenced tests and 2) to summarize how each should be used in the assessment of developmental performance in children. Specific developmental assessments, the populations they address, and the information they provide are described briefly. The need for additional criterion-referenced tests in physical therapy is discussed, and an example of how task analysis can be applied to movement or motor skills in the development of a criterion-referenced test is provided. Physical therapists can enhance the credibility of their assessments by appropriate use of norm-referenced and criterion-referenced tests.

Key Words: Pediatrics, evaluation; Physical therapy.

Physical therapists use various tests to document children's motor behavior and sensory-perceptual, cognitive-language, and social-emotional functioning. The purposes of this article are 1) to define norm-referenced and criterion-referenced tests and 2) to describe their use in pediatrics. In addition, the need for criterion-referenced tests that are related to physical therapy intervention is discussed, and an example of how task analysis can be applied to movement and used in the development of a criterion-referenced test is included.

NORM-REFERENCED AND CRITERION-REFERENCED TESTS

Physical therapists often use published developmental tests to ensure credibility in the assessment of children. The purpose of testing must be determined by the therapist if the appropriate test is to be selected. If the purpose of a test is to identify children with motor handicaps in a school system or to establish a specific age level for an individual child, a norm-referenced test would be used. If a physical therapist wants to plan an intervention program or determine its effectiveness by examining an individual child's performance, however, a criterion-referenced test would be appropriate.

Norm-referenced tests are designed to examine individual performance in relation to the performance of a representative group. Criterion-referenced tests document individual performance in relation to a domain of information or specific set of skills. A standardized test is inferred usually to be a norm-referenced test, but criterion-referenced tests also may be standardized. Standardization is the process of administering a test under uniform conditions to each child who is to be tested. A normative test is a test that has been administered in a standardized manner to large numbers of children at various ages. A test, therefore, may be standardized in its administration, but not normed, as in some criterion-referenced tests. In addition to purpose, factors such as how the test was developed and how the test scores are interpreted help to distinguish between norm-referenced and criterion-referenced tests. According to Swezey, a test is criterion-referenced if the scoring is based on absolute rather than relative standards, its primary use is to measure mastery of specific skills or tasks, and the test items are based on known performance objectives associated with the tasks of interest.

Glaser introduced the term criterion-referenced measures. He stated that the concept of proficiency measurement was a continuum of skill ranging from no proficiency at all to a perfect performance. Knowledge of an individual's performance on a criterion-referenced measure provides explicit information as to what he can and cannot do, thereby providing an index of competence that is independent of the performance of others. The child's performance on criterion-referenced tests is compared with an external criterion or standard of performance without regard to the distribution of scores achieved by other individuals of the same age. Unless the criterion-referenced test is normed and administered in a standardized manner, it cannot be used to assign age levels as a normative test. Tests that are domain referenced, competency based, and objectives referenced can be subsumed under the category of criterion-referenced tests.

Criterion-referenced tests are related directly to instructional objectives, are based on task analysis, and are designed to measure changes in successive performances of an individual. Criterion-referenced tests, therefore, are sensitive to and can be used to measure the effects of instruction. Conversely, norm-referenced tests generally are not related to instructional objectives, do not use task analysis, and are designed to delineate differences among individuals. Norm-referenced tests, therefore, are not sensitive to and should not be used to evaluate the effects of instruction.

Although some uncertainty has existed about appropriate measures for determining reliability and validity of criterion-referenced instruments, models of test theory can be applied to criterion-referenced tests. Criterion-referenced tests should demonstrate the same interrater and test-retest reliability as norm-referenced tests, and issues of validity (particularly content, concurrence, and predictive value) also are applicable, depending on the purpose of the test. Similarities and differences between norm-referenced and criterion-referenced tests are summarized in Table 1.
TABLE 1
Similarities and Differences Between Norm-Referenced and Criterion-Referenced Tests

<table>
<thead>
<tr>
<th>Norm-Referenced Tests</th>
<th>Criterion-Referenced Tests</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Purposes:</strong> to examine individual performance in relation to a representative group; can be used to establish age levels; used for diagnosis and placement.</td>
<td><strong>Purposes:</strong> to examine individual performance in relation to a criterion or external standard; cannot be used to establish age levels unless normed; used for program planning and evaluation because items are sensitive to effects of instruction (intervention).</td>
</tr>
<tr>
<td>Test construction: items usually not developed from task analysis; test items may or may not be related to the objectives of instruction (intervention).</td>
<td>Test construction: items developed from task analysis; test items are related to the objectives of instruction (intervention).</td>
</tr>
<tr>
<td>Administration: must be administered in a standard manner.</td>
<td>Administration: may or may not be administered in a standard manner.</td>
</tr>
<tr>
<td>Scoring: based on standards relative to a group; variability of scores is desired with normal distribution.</td>
<td>Scoring: based on absolute standards; variability of scores is not obtained because perfect or near-perfect scores are desired.</td>
</tr>
<tr>
<td>Psychometric properties: test should demonstrate reliability and validity.</td>
<td>Psychometric properties: test should demonstrate reliability and validity.</td>
</tr>
</tbody>
</table>

NORM-REFERENCED TESTS IN PEDIATRICS

Two examples of norm-referenced tests for young children are the Bayley Scales of Infant Development and the Gesell developmental scales. The Bayley scales were standardized on 1,400 children between 1 and 15 months of age and 160 children between 18 and 30 months of age. The main purpose of the Bayley scales is to establish current developmental status (motor and mental scales) to identify problems and the need for intervention. The Gesell scales were administered to groups of children between 4 weeks and 36 months of age with the purpose of identifying even minor deviations in the areas of adaptive, gross and fine motor, language, and personal and social development. The motor performance of older children can be evaluated using the Bruininks-Oseretsky Test of Motor Proficiency (ages 4.5-14.5 years) or the Stott-Moyes-Henderson Test of Motor Impairment (ages 5-12 years). Both tests have normative data and are designed to be used in making decisions about appropriate educational and therapeutic placement.

CRITERION-REFERENCED TESTS IN PEDIATRICS

The Brigance Diagnostic Inventory of Early Development can be used with children whose developmental ages range from birth to 7 years. This criterion-referenced test is based on items from norm-referenced tests, and its purpose is to assess general development as a guide to subsequent instruction. Developed as a response to Public Law 94-142 for children with developmental handicaps, it is used appropriately for programming, but not as a diagnostic or placement test. A second criterion-referenced test is the Milani-Compar etti Motor Development Screening Test. This test is a neurodevelopmental examination of children from 0 to 2 years of age with tasks designed to measure developmental reflexes and motor skill development. Although its purpose is to determine “whether one child’s motor development corresponds to that of a normal child,” normative data are unavailable.

Another criterion-referenced test is the Movement Assessment of Infants (MAI), which was designed to provide measures of muscle tone, reflex development, automatic movement, and volitional movement in the first year of life and to provide an assessment of risk for motor dysfunction at 4 months of age. The MAI is a good example of a criterion-referenced test that currently is being subjected to reliability and validity testing and to the process of standardization and normative data collection so that its use as a norm-referenced test eventually may be appropriate. The Peabody Developmental Motor Scales (ages 0-7 years) is an example of a norm-referenced test that also can be used as a criterion-referenced test, depending on the purpose of its use and the interpretation of the child’s performance.

USING TEST RESULTS

Although formal testing is only a part of the overall assessment of children, quantifiable data are essential for communicating information regarding motor and other types of performance and developmental levels. Norm-referenced tests, however, are not particularly helpful to the physical therapist in planning qualitative intervention and may be used inappropriately to develop intervention objectives or to teach the next developmental skill. As motor development tests are developed, items often are chosen from previously published normative tests, resulting in similar test items that are reflective of major motor milestones, rather than documenting qualitative factors that may exist. Tests measuring sensory-based functions are more innovative and often more qualitative in nature than tests of motor development. These tests may have more questionable construct validity, however, because they are more theoretical in nature. Sensory-based tests such as the Miller Assessment for Preschoolers (ages 2 years 9 months to 5 years 8 months) and the Sensory Integration and Praxis Tests (ages 5-10 years) complement tests of motor skills and can be psychometrically sound.

Physical therapists use clinical observations and expertise in assessing motor development to determine why a child has motor delays or dysfunction, and intervention usually is aimed at underlying problems. Criterion-referenced motor development tests consisting of task analyses, therefore, are usually more helpful in planning intervention programs and measuring change than are norm-referenced tests.

CONSTRUCTING CRITERION-REFERENCED TESTS

Physical therapists need more criterion-referenced tests to evaluate the sensory and motor abilities of children and adults. The characteristics of well-constructed criterion-referenced tests have been established, and physical therapists should consider constructing tests that meet their specific program...
planning and evaluation needs. Criterion-referenced tests can be designed for a particular group or subgroup of patients or to assess specific functional abilities.

In developing criterion-referenced tests, several models can be used. According to Swezey, in criterion-referenced measurement test items are developed directly from an analysis of tasks to be performed. For physical therapists working with children, these tasks often would be related to normal motor development. Current criterion-referenced tests often provide a standard of performance for motor skills, but may not use external criteria of interest to physical therapists. For example, rolling is a developmental motor skill often assessed as either present or absent. Table 2 illustrates some performance components related to physical therapy intervention that could be observed during rolling, such as the ability to roll in both directions, motor planning, the influence of postural reactions and primitive reflexes, and the vestibular response to self-induced movement. Specific standards of performance by which to judge the child’s performance are listed in the Appendix. Such a task analysis would be appropriate for use as a pretest and posttest device to identify problem areas underlying poor performance. This information would be helpful to the therapist in treatment planning and in subsequent revaluations of performance as one way to measure change. Performance on such a test could be used as a screening technique, but unless criterion-referenced tests are administered to a large number of children at different ages, they could not be used as diagnostic tools for assigning specific age (or normative) levels or for determining program placement.

**CONCLUSIONS**

Physical therapists include norm-referenced and criterion-referenced tests in the assessment process. One essential element in determining the appropriateness of the test selected is how it is to be used. Norm-referenced tests, in which the child’s performance is compared with that of a normative group, are used for diagnostic and placement purposes. Criterion-referenced tests compare a child’s performance against his own performance or an external standard, rather than against the performance of others. Criterion-referenced tests

---

**APPENDIX**

**Analysis of Rolling**

Instructions: Child rolls about 12 ft in a straight line (place masking tape line on carpet or mat). Ask child to roll from one end to the other and back. Demonstrate, if necessary.

Criteria (Score: 1 = Poor, 3 = Fair, 5 = Optimal)

**R**olls **both directions**. Poor: Child rolls to one direction only. Fair: Child tends to roll to one direction, but will roll both directions when encouraged (check side to which child does not roll spontaneously). Optimal: Child rolls both directions easily and spontaneously.

**Motor** planning. Poor: Child cannot roll from one point to another; may roll in circular pattern. Fair: Child has some difficulty rolling in a straight line, but can correct direction to arrive at designated point. Optimal: Child rolls easily in a straight line in both directions. If response is asymmetric, check side to which child has more difficulty motor planning.

**Body righting**. Poor: Child log rolls; no segmentation of hip, trunk, or shoulders. Fair: Child log rolls, but occasionally uses segmental movement. Optimal: Child performs smooth segmental roll in both directions with hip or shoulder leading. If response is asymmetric, check side to which child does not use body righting.

**Head righting**. Poor: As child rolls, head touches floor occasionally, or child cannot lift head off surface. Fair: Child lifts head up, but occasionally drops it without touching floor. Optimal: Child’s head “rights” in all positions and does not touch floor. Evaluate when child is in prone, supine, and side-lying positions during rolling.

**Asymmetrical tonic neck response (ATNR)**. Poor: Child uses ATNR to roll by turning face to elicit arm extension and increase momentum for rolling (check direction face is turned if ATNR response occurs). Fair: Child occasionally uses ATNR to roll. Optimal: Child rotates left shoulder before right shoulder when rolling to right side (if child uses ATNR, right arm will lead). Evaluate when child is rolling in both directions.

**Antigravity flexion**. Poor: Child routinely assumes hyperextended position of neck or trunk when positioned supine. Fair: Child occasionally assumes hyperextended position when positioned supine. Optimal: Child assumes extended position of trunk momentarily when positioned supine.

**Antigravity extension**. Poor: Child assumes flexed posture when in prone position (note whether hips, knees, and arms are flexed). Fair: Mild flexion is evident when in prone position. Optimal: Child does not flex when in prone position.

**Becomes dizzy** (vestibular). Poor: Child reports or displays dizziness or nausea or becomes disoriented and cannot complete task. Fair: Child reports dizziness or nausea or becomes disoriented, but can complete task with some difficulty. Optimal: Child may report mild dizziness, but does not become disoriented and completes task easily.

---

**TABLE 2**

**Analysis of Rolling**

<table>
<thead>
<tr>
<th>Task</th>
<th>Performance Level*</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Poor</td>
</tr>
<tr>
<td>Rolls both directions</td>
<td>1</td>
</tr>
<tr>
<td>Motor planning</td>
<td>1</td>
</tr>
<tr>
<td>Body righting</td>
<td>1</td>
</tr>
<tr>
<td>Head righting</td>
<td>1</td>
</tr>
<tr>
<td>Antigravity flexion</td>
<td>1</td>
</tr>
<tr>
<td>Antigravity extension</td>
<td>1</td>
</tr>
<tr>
<td>Hips flexed</td>
<td>1</td>
</tr>
<tr>
<td>Knees flexed</td>
<td>1</td>
</tr>
<tr>
<td>Arms flexed</td>
<td>1</td>
</tr>
<tr>
<td>Becomes dizzy (vestibular)</td>
<td>1</td>
</tr>
</tbody>
</table>

**TOTAL**

* No score given if subject unable to perform task. (See Appendix for criteria.)

---

*1 ft = 0.3048 m.*
are used for program planning and evaluation, and task analyses used in the development of criterion-referenced tests can be applied to movements or motor skills of interest to physical therapists. In addition, external standards of performance can be chosen that relate directly to physical therapy intervention. Although physical therapists use their observational skills to evaluate the qualitative aspects of movement, norm-referenced and criterion-referenced tests are essential for ensuring credibility in the assessment process and for communicating information regarding the need for intervention, the type of intervention necessary, and changes in the child's function. Such data are essential not only to the physical therapist, but also to the child's family, other health care professionals, and third-party payers.

REFERENCES

11. Milani-Comparetti A: The Milani-Comparetti Motor Development Screening Test. Omaha, NE, Meyer Children’s Rehabilitation Institute, University of Nebraska Medical Center, 1977
15. Ayres AJ: Sensory Integration and Praxis Tests. Pasadena, CA, Western Psychological Services, to be published