One of the major changes in the testing field over the last 20 years has been the increased interest in and use of criterion-referenced tests (CRT). Criterion-referenced tests provide a basis for assessing the performance of examinees in relation to well-defined domains of content rather than in relation to other examinees, as with norm-referenced tests. Criterion-referenced tests are now widely used (1) in the armed services, to assess the competencies of servicemen; (2) in industry, to assess the job skills of employees and to evaluate the results of training programs; (3) in the licensing and certification fields, to distinguish "masters" from "non-masters" in over 900 professions in the United States alone; and (4) in educational settings such as schools, colleges, and universities, to assess the performance levels of students on competencies of interest.

The primary purpose of this chapter is to provide an introduction to the field of criterion-referenced measurement and to several of the technical developments that provide the measurement framework for constructing, evaluating, and using these tests in a wide number of educational settings.

A Brief History and Some Central Ideas

Over 900 papers have been published on the topic since that time, and the direction of educational testing has been changed dramatically. Glaser (1963), and later Popham and Husek (1969), were interested in an approach to testing that would provide information necessary for making a variety of individual and programmatic decisions arising in connection with specific objectives, skills, or competencies. The Popham-Husek article was especially effective in stimulating educational measurement specialists to take up the challenges associated with criterion-referenced measurement. Norm-referenced tests were regarded as limited because they could not provide the desired kinds of information.

Standard procedures for testing and measurement within a norm-referenced framework are well known to psychologists, but these procedures are far less appropriate when the questions being asked concern what examinees can and cannot do or what levels of proficiency examinees have attained in relation to a set of objectives or skills of interest (Glaser, 1963; Hambleton & Novick, 1973; Popham & Husek, 1969). Norm-referenced tests are constructed principally to facilitate the comparison of individuals (or groups) with one another or with respect to a norm group on the trait measured by the test. In contrast, criterion-referenced tests (or, as they are sometimes called, “proficiency tests,” “domain-referenced tests,” “mastery tests,” “competency tests,” “credentialing exams,” or “basic skills tests”) are constructed to permit the interpretation of individual (and group) test scores in relation to a set of clearly defined objectives, skills, or competencies.

Many definitions of criterion-referenced tests have been proposed in the literature over the last 20 years (see, for example, Nitko, 1980; Popham, 1978a). In fact, Gray (1978) reported the existence of 57 different definitions! At present the most widely accepted definition comes from Popham (1978a): “A criterion-referenced test is used to ascertain an individual’s status with respect to a well-defined behavioral domain” (p. 93). Four points about the definition are worth noting. First, terms such as objectives, skills, and competencies are now used interchangeably in the measurement literature and by persons working in the testing field. Second, each objective measured in a criterion-referenced test must be well defined. The emphasis for this requirement goes back to work by Popham (1974). Well-defined objectives facilitate the process of writing test items and enhance the validity of test score interpretations. Item writing is improved because well-defined objectives provide a framework within which item writers and item reviewers can work. Validity is enhanced because of the clarity of the content or behavior domains to which test scores are referenced. The breadth and complexity of each