Beginning in the fall of 1979, at their respective institutions, each author used criterion-referenced measurement to assess student achievement in the junior- and senior-level occupational therapy course in psychosocial dysfunction. Based on their experience, the authors offer a guide for the construction of criterion-referenced test items and discuss the implementation of criterion-referenced measurement in a traditional college setting.

Since the mid-1960s there have been major movements in clinical practice in health care, in the profession of occupational therapy, and in allied health-care education. Health care professionals have shown increased concern for demonstrating visibly the outcomes of their intervention. For occupational therapists this concern has been reflected in clinical practice in steps taken to clearly identify changes in client behavior vis-a-vis behavioral goal writing and the problem-oriented record. Concurrently, the profession has initiated studies to determine performance competencies for the entry-level therapist and to specify criteria for certification, licensure, and standards of practice. At the same time in education, academicians began writing classroom instructional objectives that reflected directly the criterion competencies necessary for occupational therapy students. Thus competency-based curricula emerged that emphasize mastery learning and the evaluation of student competence by means of criterion-referenced measurement.

Survey of the Literature
Curriculum models and design (1-3), competency education (4-7), the certification examination (8-12), and mechanisms for competency assurance (13-15), have been discussed previously in occupational therapy literature, as well as in multiple allied health publications.

Curricula previously designed according to a subject-centered or an integrated model have moved toward a competency-based design. A competency-based educational program has been described as one “with a curriculum organized around functions or competencies required for the practice of a medical profession in a specified setting. The intended outcome of a competency-based program is a health professional who can practice at a defined level of proficiency, in accordance with local conditions, to meet local needs.” (16, p 18)

From the literature, an historical perspective on competency-based education in occupational therapy can be outlined. The movement that was conceived in the 1960s yielded tangible outcomes in the 1970s. In 1972, the American Occupational Therapy Association (AOTA), funded by the Bureau of Health Manpower Education in the National Institutes of Health (NIH), initiated a project that produced The Roles and Functions of Occupational Therapy Personnel, a task-oriented delineation of the skills seen as comprising the roles of the OTR and COTA (17). In 1974, the AOTA received NIH funds for the development of A Curriculum Guide for Occupational Therapy Educators (18). A contract awarded in 1974 yielded standards...
of practice and information that could guide self- and peer review and develop continuing education programs (19). During the years 1976 to the present time, we have witnessed further role studies by the AOTA (20-22).

This article focuses on unpublished information describing the application of criterion-referenced measurement systems in the college classroom. Since competency-based education and related terminology have been defined, but with some variance, key terms will be reviewed. Also, the use of these terms in the discussion of developing a criterion-referenced measurement system will be clarified. Then criterion-referenced test construction and a model for classroom implementation are discussed, as well as the advantages and disadvantages of implementing criterion-referenced measurement.

Definitions of Concepts
The term competency is the key in the context of competency-based education. Competencies, as defined by Gilfoyle, are, "general statements describing the skills, knowledge and attitudes needed for practice" (23, p 2). These statements "communicate a level of generality which indicates what the graduate will be able to do" and are "based upon identified roles and functions for performance expectations." (23, p 2). Each competency statement can be further broken down into specific statements of behavior that indicate "what" the entry-level therapist will be able to do. The specific behaviors that must be mastered for each competency to be achieved are known as the behavioral criteria (23, p 3).

Since mastery of given criteria is desirable in a competency-based program, the term mastery learning is sometimes used to describe this educational approach. The term mastery implies that "given adequate preparation, unambiguous learning goals, sufficient learning resources and a flexible time schedule, the student will achieve a defined competency at a high level of proficiency" (16, p 19). It is important that each instructor and each curriculum establish what level of performance constitutes a "high level of proficiency." Setting performance standards must depend on desired terminal outcomes (behaviors seen as necessary for professional entry), and also on the relation of a particular course to a whole curriculum. For example, one could cite art media as a therapeutic modality. The sophomore student might be expected to discuss the uses of art as a vehicle to promote emotional expression and interpersonal functioning, whereas the senior might be expected to demonstrate the use of art media in a small group treatment simulation.

The traditional approach to measurement of learning in the college classroom is norm-referenced measurement. The norm-referenced test measures a student's status in relation to the performance of other students in a group who also completed the tests." (24, p 24) Instructors who expect that in their course there will be roughly 10 percent As, 30 percent Bs, 40 percent Cs, and so on, will likely follow a norm-referenced approach. This instructors' goal when designing a testing experience is frequently to differentiate the few "excellent" students as well as to indicate the few students who will be expected to fail. Norm-referenced measurement is not the preferred assessment of achievement in a competency-based program, but it may be preferred where testing is designed to limit and/or order students, as in a selection process.

Criterion-referenced testing is the identified measurement of learning in competency-based education. A criterion-referenced test is used "to ascertain an individual's status with respect to a well-defined performance objective," or performance criterion (24, p 14). Well-defined means that the conditions under which the criterion behavior will occur and the degree of proficiency of the performance are explicit. In practical application, judgment is exercised so that the student performance objectives and test items are clear, meaningful, and not so narrowly defined as to limit their applicability.

Often, criterion-referenced tests are used initially, via a pre-test, to ascertain the knowledge and skills the student brings into a course. These tests are used throughout the duration of a course to determine whether students have achieved mastery of new topics covered. A test given after a learning module or unit is known as a post-test. Students who demonstrate that they have not mastered specified criterion behaviors may be guided to remedial or alternative learning experiences, then retested. Thus, criterion-referenced testing is sometimes referred to as mastery testing, although the two terms are not identical. The instructor who designs a criterion-referenced test should not discriminate among those students who have participated in a specified learning experience; rather, the test should discriminate between those students who have participated in a given learning experience and those who have not.

Many colleges and universities use written student performance objectives in establishing course goals. A student performance objective, frequently referred to as a be-
Behavioral objective, is a statement that describes what a student will be able to do after a learning experience. In competency-based education, classroom performance objectives should lead to or reflect the criteria comprising entry-level competencies.

**Constructing a Criterion-Referenced Test**

Since the criteria leading to professional competency reflect all three of Bloom's learning domains (25), it must be first emphasized that criterion-referenced testing is not exclusively paper and pencil testing. Especially where the affective and psychomotor domains of learning are concerned, the use of classroom performance testing (demonstration) is an excellent vehicle for measuring learning. However, in this article the focus is on the design of written tests.

When constructing a criterion-referenced test, it is most important to realize that a test item need not look different from a norm-referenced test item. However, each item must clearly relate to a performance objective, and this objective must reflect performance criteria that lead to professional competency.

A second key factor in designing a criterion-referenced test lies in the instructor’s attitude or intent in writing the test. Rather than trying to construct a test that will “catch” all but the best students, the instructor is seeking to develop test items that are most representative of the material covered. The following are guidelines for constructing and evaluating a criterion-referenced measurement instrument:

1. Behavioral performance objectives are identified and performance standards established. The instructor must define behaviors the student will demonstrate in this course and the degree of accuracy in performance that will constitute a passing grade (or mastery, if mastery is the goal). These target behaviors and standards may be established by accumulating data on typical student performance; by using the combined judgment of faculty and students to determine reasonable standards; and by looking to clinicians for guidelines regarding performance levels seen as essential to beginning practice (24, pp 165-168). Regardless of the approach, there must be a basis for establishing performance objectives and performance standards.

2. Good criterion-referenced tests can be constructed only if performance objectives are explicit and are stated in measurable terms. These objectives are made available to the student.

3. Test items are worded or designed to clearly test what the author purports to measure, and under the conditions specified in the performance objectives.

4. The instructor should make no special attempt to increase the allure of wrong answers.

5. All performance objectives in each of the learning domains should ultimately be tested. One test item may cover more than one objective; conversely, more than one test item may be needed to test completely one objective (24, pp 99-100).

6. In the test design, the proportion of items devoted to specific areas of content and the levels and domains of learning reflect the instructor's decisions regarding what is the most important course content and the class instructional design.

7. Content or "descriptive" validity is essential; that is, the test measures what it contends to measure (24, p 156).

8. The test is reliable. One method of establishing reliability may be established by the equivalent-and-stability approach also used in the norm-referenced testing system; that is, two forms of the test are created by selecting items from a pool of homogeneous test items, then, the first version is given, followed by administration of the second version to the same students at a later time (24, p 154). The test is evaluated for its reliability in discriminating between those who have and those who have not completed a given learning experience. To that end, one would expect a variance in the scores of the two groups (26, p 86).

9. Homogeneity and descriptive validity of test items are judged by content specialists, such as practicing therapists. If practicing therapists are not easily available, faculty colleagues can be helpful in reading tests for content validity.

10. Uniformly high scores on a criterion-referenced test do not necessarily mean that the test was poorly constructed, but rather, may indicate that performance criteria were clear and attained. To assure that high test scores indicate behavioral change that resulted from participation in a given learning experience, the instructor might give the test to comparable students who had not participated in the learning experience. High scores would not be expected.

**Determine Competency Level**

Criterion-referenced tests are frequently used to determine college grades, with grading often a sensitive and critical issue. Although educational settings handle grading differently, one option is to set a "desired" rather than a "minimal" performance standard by establishing a "B" or "A" as the passing grade. The "B" grade corresponds to 80-89
percent accuracy, the "A" to 90-100 percent accuracy. Students who perform at less than the "B" (or "A") level are given remedial learning experiences and are asked to take equivalent alternate examinations until the desired performance is demonstrated. The highest score attained is ultimately recorded for the student.

In settings where a "C" grade is considered passing, the instructor may feel pressure to set "C" and 70 percent accuracy as the passing grade, but may still encourage students to work toward a "B" or "A."

The number of times tests may be repeated and the time allowed for the completion of a learning module often depends on the flexibility of the overall curriculum design.

Figure 1
The progression from professional competency to student performance objectives

PROFESSIONAL COMPETENCY AREA: Report and Discuss
"The scope of the OTR's role in giving and receiving written and oral information about a client includes descriptions of the client's observable behaviors and analyses of the components of performance which enhance or impair client's occupational performance. To participate in such interaction with client, family or with other staff the OTR must." (22, p. 13)

PROFESSIONAL COMPETENCY: "2. explain how given clinical conditions may affect client's occupational performance" (22, p. 13)

LEARNING MODULE: FREUDIAN ANALYTIC TREATMENT

STUDENT COMPETENCY: Upon completion of class instruction, discussion and assigned readings, the student will demonstrate the roles and interrelationships of intrapsychic components in given case illustrations, using the Freudian analytic

STUDENT OBJECTIVE: "Ms. L," "Mr. C," and "Mindy"

—Case assignments

Class discussion

Exam I

First, recently defined professional competencies of the entry-level therapist were reviewed (22). One such "professional competency" is cited in Figure 1. Clarification of professional competencies can be supplemented and updated annually through discussions with and the solicitation of feedback from practitioners, practicum Field Work I and II supervisors and students; through graduate surveys; and through the exchange of information at the AOTA Council on Education meetings.

These professional competencies were then divided into criteria: statements that specified knowledge, skills, and beliefs as behavioral
components of professional competence. Which criteria and to what extent they could logically be taught in the course was determined with faculty at faculty retreats, where all courses and course content could be coordinated.

After determining target outcomes for behavioral criteria, the authors organized the criteria according to logical learning units, called learning modules. In Figure 1, the learning module is "Freudian Analytic Treatment."

Next, the content of the module was organized according to broad competency statements specifying components of the practitioners' roles and functions that could reasonably be attained by senior students. These components are called student competencies (see Figure 1).

The authors then identified specific performance objectives, each of which was seen as comprising or leading to a specific student competency (Figure 1).

The authors next designed module learning activities and determined the time frame in which they would be covered. Activities such as lectures, demonstrations, videotapes, case simulations, course handouts, and test items were all designed to focus on the achievement of competencies through the participation of the learner, with individualized opportunities to achieve "mastery" of criteria.

Both when preparing daily lesson plans, and after lecture and laboratory experiences, the authors wrote test items designed to test the performance objectives that had been covered. Test items were written weekly on individual index cards, which were then coded according to the performance objective being tested, the learning domain being tapped and the level of difficulty of the test item (see Figure 2). Since it had been decided to allow students who had not achieved a "B" to repeat examinations following remedial work, three equivalent test items were written concurrently, thus making the formation of alternate, equivalent examinations more expedient.

Upon entering the course, the students were given a written description of the course content, organized according to specified modules; a delineation of student competencies and their "enabling" student performance objectives. Students were encouraged to use these handouts as a guide for organizing their notes and to prepare for examinations.

Students who entered the course were told that they would have two to three opportunities (depending on the learning assignment) to retake examinations or do remedial assignments if they did not achieve a grade they considered satisfactory. Because a grade of "C" was established as a passing grade at both of the occupational therapy programs where this was implemented, the authors could not insist that students redo any of their work. However, almost all of the students who received less than an "A" grade opted for alternate equivalent work.

Students were generally allowed 7 to 10 days in which to repeat the assignment or to take an alternate exam. Although not ideal, students were expected to keep up with new material, even while working on remedial assignments. To stay within the time constraints set by the curriculum, students were expected to finish the course by semester's end.

To improve the descriptive validity of test items on examinations, students were encouraged to critique all examinations following grading. Also, other faculty members reviewed course examinations. While statistical validity has yet to be established, students consistently remarked that they found the examinations "straightforward," "clear," and "fair," and commented that the tests were challenging but not "intimidating."

Advantages of Criterion-Referenced Measurement

Criterion-referenced measurement is particularly advantageous during instruction designed to prepare the student for entry into a service profession, such as occupational therapy, where public welfare demands that only a qualified practitioner be involved in providing service (27, p
As described earlier, there is a close bond between mastery or competency-based learning and the evaluation of achievement by means of criterion-referenced measures. With this in mind, the advantages of criterion-referenced testing will be discussed with the following elements included: test items will accurately reflect student performance objectives; student performance objectives will accurately reflect the performance criteria necessary for professional entry; students who do not meet the established level of performance competence will be expected to do remedial work and be retested until the desired level of performance is achieved.

One can first delineate advantages that are tied directly to allow students the opportunity to achieve mastery. While further study is needed to establish the effect of mastery testing, some educators (28-30) have proposed the following advantages:

1. Students are more likely to conclude a learning experience feeling they have succeeded. Where traditionally only the few "best" students felt good about their own performance and received the coveted "A," this new method of evaluation provides a mechanism for most students to achieve success.

2. Competition within a class is diminished. Students compete mostly with themselves to improve their own performance.

3. Performance anxiety is diminished.

4. Students feel the support of their instructor as they work together to enhance competence level.

5. The use of feedback and individualized alternative learning experiences is encouraged.

6. The profession can be more confident that all prospective practitioners have achieved the necessary level of competence.

Other advantages relate to the design of the tests. These are discussed in Popham's book (31), and include:

1. Students are clear about what is expected of them.

2. Organization of material is enhanced. Planning instruction around performance criteria promotes a careful selection of material by the instructor, and helps to highlight important points. Similarly, there is meaningful structure by which students can organize their study.

3. The relevance of classroom instruction to practice is evident, as is the relationship of test items to course objectives.

4. The teaching-learning process is demystified. Students, more cognizant of what is expected, then do not go into an examination wondering if they are prepared; nor, at the conclusion of an examination must they wonder about the areas that need remediation.

5. Students are more able to be independent in pursuit of learning. With demystified course requirements, students can take more responsibility for their own growth.

6. Instructors have clear and absolute information about what their students can and cannot do.

7. Change, or improvement toward a certain level of competence, can be evaluated for each student.

In summary, the advantages to a profession using criterion-referenced measurement in evaluating achievement toward a given competency base lies in establishing a logical and discernable relationship between professional entry-level expectations and classroom learning, in providing some absolute information about student movement toward competency, and in assuring a certain level of performance in prospective practitioners. Increased student confidence and increased student independence are related benefits that can ultimately serve to strengthen the profession.

### Disadvantages

An obvious disadvantage of criterion-referenced measurement is the time and work needed to initiate it. The establishing of appropriate classroom performance objectives consonant with professional entry-level competencies is a large task. Agreement among instructors about what constitutes competency in a given course, and how competency can be tested, may be difficult to achieve. Having written behavioral course objectives does not ensure that tests are criterion-referenced, and tests may need to be modified.

In evaluating the use of criterion-referenced measurement where competency at a specified level is the goal, ask the following questions:

1. Will the extra time needed for paperwork and alternate assignments benefit the student?

2. How many times should a student repeat an assignment or exam, when a course must be completed within school calendar boundaries?

3. Will students feel as good about an "A" grade if an "A" is available to most of the class?

4. Will it be difficult for students to become accustomed to their new roles by using this nontraditional measurement system?

5. In encouraging performance to specified levels, is student exploration beyond required class tasks discouraged?

6. In encouraging performance to specified levels on given criteria, are thinking and creative problem-solving discouraged?
7. Will clarity in defining performance objectives lead to an emphasis on triviality by both the instructor and the student, rather than an emphasis on concept formation?

8. Are mediocre students rewarded more than excellent students?

9. How will the effectiveness of this new testing approach be evaluated?

10. Does decreased anxiety facilitate learning?

Summary
One format was presented that met the practical parameters of classroom education while improving the clarity and effectiveness of classroom testing, and facilitating the assurance of future practitioner competence. While employing criterion-referenced testing, this format did not adhere strictly to a competency-based model, nor were test questions evaluated for statistical validity or reliability.

Criterion-referenced measurement has the potential to be a meaningful tool in professional education, but it is not a panacea for educational demands. Difficult questions confront every educator who considers using criterion-referenced measurement toward a goal of student mastery. Criterion-referenced measurement is practical, and the educators who use it make a statement regarding their beliefs about the education process.

Acknowledgment
This article and criterion-referenced model were presented and discussed at the 1980 AOTA Annual Conference, Denver, Colorado.

REFERENCES
22. Continuation of Study of Measures of Proficiency in Occupational Therapy, Contract #231-76-0052, Phase II Written Examination and Phase III Performance Examination, 1979, American Occupational Therapy Association, Inc., Rockville, Maryland
23. Occupational Therapy in Educational Management. Module One of a Competency-Based Educational Program, 1979, American Occupational Therapy Association, Inc., Rockville, Maryland

The American Journal of Occupational Therapy 327