DOCUMENT RESUME

ED 132 411 08 CE 009 202

TITLE	Vocational Education Curriculum Specialist (VECS). Module 9: Testing Instructional Objectives. Study Guide. (Teaching/Learning Module).
INSTITUTION	American Institutes for Research in the Behavioral Sciences, Palo Alto, Calif.
SPONS AGENCY	Office of Education (DHEW), Washington, D.C.
PUB DATE	76
CONTRACT	OEC-0-74-9286
NOTE	99p.; For related documents see CE 009 129-136 and CE 009 192-206
EDRS PRICE	MF-$0.83 HC-$4.67 Plus Postage.
DESCRIPTORS	Achievement Tests; Behavioral Objectives; *Criterion Referenced Tests; *Curriculum Development; *Educational Assessment; *Educational Objectives; Evaluation Methods; Instructional Materials; Instructional Programs; Learning Modules; Learning Processes; Measurement Techniques; Post Secondary Education; Program Evaluation; Secondary Education; Specialists; Student Evaluation; Study Guides; Teacher Education; Test Construction; Testing; *Vocational Education

ABSTRACT

One of 15 core modules in a 22-module series designed to train vocational education curriculum specialists (VECS), this guide is intended for use by both instructor and student in a variety of education environments, including independent study, team teaching, seminars, and workshops, as well as in more conventional classroom settings. The guide has five major sections. Part I, Organization and Administration, contains an overview and rationale, educational goals and performance objectives, recommended learning materials, and suggested reference materials. Part II, Content and Study Activities, contains the content outline arranged by goals. Study activities for each goal and its corresponding objectives follow each section of the content outline. Content focus is on the concept of criterion-referenced measurement within the framework of educational evaluation, selecting approaches/techniques for assessing student achievement of instructional objectives in the three domains of learning, and developing an evaluation plan and constructing test instruments for measuring student achievement of instructional objectives. Part III, Group and Classroom Activities, suggests classroom or group activities and discussions keyed to specific content in the outline and to specific materials in the list of references. Part IV, Student Self-Check, contains questions directly related to the goals and objectives of the module, which may be used as a pretest or posttest. Part V, Appendix, contains suggested responses to the study activities from part II and responses to the student self-checks. (HD)

-Study Guide-

Module 9

TESTING INSTRUCTIONAL OBJECTIVES

This document is one of a series of teaching/learning modules designed to train Vocational Education Curriculum Specialists. The titles of all individually available documents in this series appear below:

INTRODUCTORY MODULES

1. The Scope of Vocational Education
2. Roles of Vocational Educators in Curriculum Management
3. Current Trends in Vocational Education
4. Organization of Vocational Education
5. Legislative Mandates for Vocational Education
6. The Preparation of Vocational Educators

CORE MODULES

1. Important Differences Among Learners
2. Learning Processes and Outcomes
3. Applying Knowledge of Learning Processes and Outcomes to Instruction
4. Assessing Manpower Needs and Supply in Vocational Education
5. Laying the Groundwork for Vocational Education Curriculum Design
6. Selecting Instructional Strategies for Vocational Education
7. Derivation and Specification of Instructional Objectives
8. Development of Instructional Materials
9. Testing Instructional Objectives
10. Fiscal Management of Vocational Education Programs
11. Introducing and Maintaining Innovation
12. Managing Vocational Education Programs
13. Basic Concepts in Educational Evaluation
14. General Methods and Techniques of Educational Evaluation
15. Procedures for Conducting Evaluations of Vocational Education

SEMINARS AND FIELD EXPERIENCE MODULE

(Seminars in Authority Roles and the Curriculum Specialist in Vocational Education, and Leadership Styles and Functions of the Curriculum Specialist in Vocational Education; field work in Project Design and Administration, Operation of School Programs, Evaluation of School Programs, Educational Research and Development, and State, Regional, and Federal Program Supervision)

INSTALLATION GUIDE

For sale by the Superintendent of Documents
U. S. Government Printing Office
Washington, D.C. 20402 Price $_ _ _
Stock No.

PREFACE

Who is a vocational education curriculum specialist? The answer to this question is not as simple as it might appear. A vocational education curriculum specialist is likely to work in many different capacities, including, but not limited to: instructor, department chairperson, dean of vocational-technical education, vocational supervisor, principal, state or local director of vocational education, and curriculum coordinator.

The specialist is, perhaps, more identifiable by his/her responsibilities, which include, but are not limited to:

- planning, organizing, actualizing, and controlling the work of an educational team performed to determine and achieve objectives.
- planning, organizing, and evaluating content and learning processes into sequential activities that facilitate the achievement of objectives.
- diagnosing present and projected training needs of business, industry, educational institutions, and the learner.
- knowing, comparing, and analyzing different theories of curriculum development, management, and evaluation and adapting them for use in vocational-technical education.

This teaching/learning module is part of a set of materials representing a comprehensive curriculum development project dealing with the training of vocational education curriculum specialists. The purpose of this two-year project was 1) to design, develop, and evaluate an advanced-level training program, with necessary instructional materials based on identified vocational education curriculum specialist competencies, and 2) to create an installation guide to assist instructors and administrators in the implementation process.

The curriculum presented here is, above all else, designed for flexible installation. These materials are not meant to be used only in the manner of an ordinary textbook. The materials can be used effectively by both instructor and student in a variety of educational environments, including independent study, team teaching, seminars, and workshops, as well as in more conventional classroom settings.

Dr. James A. Dunn
Principal Investigator and
presently Director,
Developmental Systems Group
American Institutes for Research

ACKNOWLEDGEMENTS

The Vocational Education Curriculum Specialist Project was a comprehensive development and evaluation effort involving the contribution of a large number of people: project staff, curriculum consultants, a national advisory panel, and a number of cooperating colleges and universities. This wide variety of valuable inputs makes it difficult to accurately credit ideas, techniques, suggestions, and contributions to their originators.

The members of the National Advisory Panel, listed below, were most helpful in their advice, suggestions, and criticisms.

Myron Blee	*Florida State Department of Education*
James L. Blue	*RCU Director, Olympia, Washington*
Ralph C. Bohn	*San Jose State University*
Ken Edwards	*International Brotherhood of Electrical Workers*
Mary Ellis	*President, American Vocational Association*
George McCabe	*Program Director, Consortium of California State University and Colleges*
Curtis Henson	*Atlanta Independent School District, Georgia*
Ben Hirst	*Director, Consortium of the States, Atlanta, Georgia*
Joseph Julianelle	*U. S. Department of Labor*
Lee Knack	*Industrial Relations Director, Morrison-Knudsen, Inc.*
Bette LaChapelle	*Wayne State University*
Jerome Moss, Jr.	*University of Minnesota*
Frank Pratzner	*CVE, Ohio State University*
Rita Richey	*Wayne State University*
Bryl R. Shoemaker	*Ohio State Department of Education*
William Stevenson	*Oklahoma State Department of Education*

The project would not have been possible without the cooperation and commitment of the field test institutions listed below.

California State University, Long Beach
California Polytechnic State University, San Luis Obispo
Consortium of California State University and Colleges
- California State University, Sacramento
- California State University, San Diego
- California State University, San Francisco
- California State University, San Jose
- California State University, Los Angeles

Iowa State University
University of California Los Angeles
University of Northern Colorado

Overall responsibility for the direction and quality of the project rested with James A. Dunn, Principal Investigator. Project management, supervision, and coordination were under the direction of John E. Bowers, Project Director.

TABLE OF CONTENTS

Part I:

Organization and Administration

PART I
ORGANIZATION AND ADMINISTRATION

Guidelines

This study guide has five major sections. Each section contains useful information, suggestions, and/or activities that assist in the achievement of the competencies of a Vocational Education Curriculum Specialist. Each major section is briefly described below.

PART I: ORGANIZATION AND ADMINISTRATION

PART I contains an Overview and Rationale, Educational Goals and Performance Objectives, Recommended Learning Materials, and Suggested Reference Materials. This section will help the user answer the following questions:

- How is the module organized?
- What is the educational purpose of the module?
- What specifically should the user learn from this module?
- What are the specific competencies emphasized in this module?
- What learning materials are necessary?
- What related reference materials would be helpful?

PART II: CONTENT AND STUDY ACTIVITIES

Part II contains the content outline arranged by goals. The outline is a synthesis of information from many sources related to the major topics (goals and objectives) of the module. Study activities for each goal and its corresponding objectives follow each section of the content outline, allowing students to complete the exercises related to Goal 1 before going on to Goal 2.

PART III: GROUP AND CLASSROOM ACTIVITIES

The "Activities-Resources" column in the content outline contains references to classroom or group activities and discussion questions related to specific content in the outline. These activities and discussion questions

are located in PART III and for optional use of either the instructor or the student. Both the classroom activities and discussion questions are accompanied by suggested responses for use as helpful examples only--they do not represent conclusive answers to the problems and issues addressed. Also contained in the "Activities-Resources" column are the reference numbers of the resources used to develop the content outline. These reference numbers correspond to the numbers of the Suggested Reference Materials in PART I.

PART IV: STUDENT SELF-CHECK

PART IV contains questions directly related to the goals and objectives of the module. The self-check may be used as a pre-test or as a post-test, or as a periodic self-check for students in determining their own progress throughout the module.

PART V: APPENDICES

Appendix A contains responses to the Study Activities from PART II, and Appendix B contains responses to the Student Self-Check. The responses provide immediate feedback to the user and allow the module to be used more effectively for individualized study. They have been included in the last part of the module as appendices to facilitate their removal should the user wish to use them at a later time rather than concurrently with the rest of the module.

Approximately 30 hours of out-of-class study will be necessary to complete this module.

Overview and Rationale

The purpose of this module is to provide the future curriculum specialist with the knowledge and skills to develop test instruments that measure student achievement. Many texts enumerate the purposes of a testing program. Among these are the improvement of training or instruction, the motivation of students, determination of grades, and use as a basis for selection and

guidance. For purposes of this module, however, "testing" refers specifically to the assessment of student accomplishment of the instructional objectives of a course (or instructional unit) as specified in the criteria [illegible] the term "criterion-referenced testing."

The traditional and primary form of testing in vocational education has been "norm-referenced testing." And although norm-referenced testing continues to be a viable form of testing when used for appropriate purposes, it is the criterion-referenced test that is actually the most appropriate measure of whether or not an instructional objective has been achieved. It is with this latter form of testing that this module is concerned.

The module begins by examining the concept of criterion-referenced testing within the framework of educational evaluation. Although criterion-referenced testing in the strictest sense is "measurement" and not "evaluation," in a broader sense it cannot be isolated from educational evaluation, which relies on numerous measurements to determine the merit of various educational phenomena.

Next the module examines techniques and approaches appropriate for assessing students' achievement of instructional objectives in the three domains of learning: cognitive, affective, and psychomotor. Instructional objectives are amenable to a wide variety of assessment techniques, not just the all-too-often used paper-and-pencil test. The important point is that the technique selected match the requirements of the objective. Paper-and-pencil tests, for example, are not appropriate for determining whether or not a student is able to operate a wood lathe properly.

The module then presents the curriculum specialist with a technique for developing a plan from which to construct criterion-referenced test instruments, and finally provides the specialist with an opportunity to actually construct these instruments.

This module completes the series of three modules on the development of instruction for vocational education. Module 7, Derivation and Specification of Instructional Objectives, discussed procedures both for identifying possible objectives for instruction and for writing such objectives. Module 8, Development of Instructional Materials, described the process of developing instruction to accomplish specific objectives. Now this module, Module 9, completes the picture by presenting means of assessing student achievement of the objectives of instruction.

A variety of approaches to instructional development are in practice in vocational education today. These approaches include: the integrated approach; the occupational or job analysis approach; the clusters, families, or common elements of occupations approach; the functions of industry approach; and the concept approach. (Each of these approaches is briefly described in Introductory Module 2: Roles of Vocational Educators in Curriculum Management.) This series of modules on instructional development for vocational education follows an occupational or job analysis approach because it is the most common and is often used in combination with other curriculum techniques.

Goals and Objectives

Upon completion of this module, the student will be able to achieve the following goals and objectives:

GOAL 9.1: UNDERSTAND THE CONCEPT OF CRITERION-REFERENCED MEASUREMENT WITHIN THE FRAMEWORK OF EDUCATIONAL EVALUATION.

Objective 9.11 Define the following terms: educational evaluation, educational measurement, criterion-referenced testing, and norm-referenced testing.

Objective 9.12 Identify the historical conditions that gave impetus to the use of criterion-referenced measurement.

Objective 9.13 Given a specific characteristic, determine whether that characteristic describes a norm-referenced test or a criterion-referenced test.

Objective 9.14 Distinguish between norm-referenced measurement and criterion-referenced measurement on the basis of: variability, item construction, reliability, validity, item analysis, and reporting and interpretation.

GOAL 9.2: SELECT APPROACHES/TECHNIQUES FOR ASSESSING STUDENT ACHIEVEMENT OF INSTRUCTIONAL OBJECTIVES IN THE THREE DOMAINS OF LEARNING.

Objective 9.21 Recognize appropriate techniques for assessing student achievement of instructional objectives in the cognitive, affective, and psychomotor domains.

Objective 9.22 Identify the two basic types of written test questions and describe the advantages and limitations of each.

Objective 9.23 Define the term "performance test."

Objective 9.24 Select approaches/techniques for assessing student achievement of instructional objectives of a given unit of instruction.

GOAL 9.3: DEVELOP AN EVALUATION PLAN AND CONSTRUCT TEST INSTRUMENTS FOR MEASURING STUDENT ACHIEVEMENT OF INSTRUCTIONAL OBJECTIVES.

Objective 9.31 Given an instructional objective stated in behavioral terms and a list of possible test items, identify those test items that would be appropriate for assessing the objective.

Objective 9.32 Develop an evaluation plan for assessing student achievement of the instructional objectives for a given unit of instruction.

Objective 9.33 Construct test instruments for assessing student achievement of the instructional objectives for a given unit of instruction.

Recommended Materials

Popham, W. James, ed. Criterion-Referenced Measurement: An Introduction. Englewood Cliffs, New Jersey: Educational Technology Publications, 1971.

Mager, Robert F., and Beach, Kenneth M., Jr. Developing Vocational Instruction. Belmont, California: Fearon Publishers, 1967.

Weber, Larry J., and Lucas, Stephen R. "Evaluating Student Progress." In The Individual and His Education, edited by Alfred H. Krebs. Second Yearbook of the American Vocational Association. Washington, D. C.: AVA, 1972.

Completed Study Guide for Module 7: Derivation and Specification of Instructional Objectives.

Completed Study Guide for Module 8: Development of Instructional Materials.

Suggested References

1. Armstrong, Robert J.; Cornell, Terry D.; Kraner, Robert E.; and Roberson, E. Wayne. The Development and Evaluation of Behavioral Objectives. Worthington, Ohio: Charles A. Jones Publishing Company, 1970.

2. Baker, Eva L., and Popham, W. James. Expanding Dimensions of Instructional Objectives. Englewood Cliffs, New Jersey: Prentice-Hall, Inc., 1973.

3. Boyd, Joseph L., Jr., and Shimberg, Benjamin. Directory of Achievement Tests for Occupational Education. Princeton, New Jersey: Educational Testing Service, 1971.

4. Boyd, Joseph L., Jr., and Shimberg, Benjamin. Handbook of Performance Testing: A Practical Guide for Test Makers. Princeton, New Jersey: Educational Testing Service, 1971.

5. Butler, F. Coit. Instructional Systems Development for Vocational and Technical Training. Englewood Cliffs, New Jersey: Educational Technology Publications, 1972.

6. Byers, Edward E., and Huffman, Harry H. Writing Performance Goals: Strategy and Prototypes. New York: Gregg Division, McGraw-Hill Book Company, 1971.

7. Clark, D. Cecil. Using Instructional Objectives in Teaching. Glenview, Illinois: Scott, Foresman and Company, 1972.

8. Cross, Aleene. Home Economics Evaluation. Columbus, Ohio: Charles E. Merrill Publishing Company, 1973.

9. Designing Objective, Essay, and Performance Tests. College Station, Texas: Vocational Industrial Education, Texas A & M University, 1970.

10. Drawbaugh, Charles C., and Hull, William L. Agricultural Education: Approaches to Learning and Teaching. Columbus, Ohio: Charles E. Merrill Publishing Company, 1971.

11. Glaser, Robert. "A Criterion-Referenced Test." In Criterion-Referenced Measurement: An Introduction, edited by W. James Popham. Englewood Cliffs, New Jersey: Educational Technology Publications, 1971.

12. Glaser, Robert. "Instructional Technology and the Measurement of Learning Outcomes: Some Questions." American Psychologist 18 (1963).

13. Goldstein, Irwin L. Training: Program Development and Evaluation. Monterey, California: Brooks/Cole Publishing Company, 1974.

14. Kibler, Robert J.; Cegala, Donald J.; Barker, Larry L.; and Miles, David T. Objectives for Instruction and Evaluation. Boston: Allyn and Bacon, 1974.

15. Mager, Robert F. Goal Analysis. Belmont, California: Fearon Publishers, 1972.

16. Mager, Robert F., and Beach, Kenneth M., Jr. Developing Vocational Instruction. Belmont, California: Fearon Publishers, 1967.

17. Miller, W. R., and Rose, H. C. Instructors and Their Jobs. 3rd ed. Chicago: American Technical Society, 1975.

18. Pautler, Albert J. Teaching Shop and Laboratory Subjects. Columbus, Ohio: Charles E. Merrill Publishing Company, 1971.

19. Popham, W. James. Evaluating Instruction. Englewood Cliffs, New Jersey: Prentice-Hall, Inc., 1973.

20. Popham, W. James, ed. Criterion-Referenced Measurement: An Introduction. Englewood Cliffs, New Jersey: Educational Technology Publications, 1971.

21. Pucel, David J., and Knaak, William C. Individualizing Vocational and Technical Instruction. Columbus, Ohio: Charles E. Merrill Publishing Company, 1975.

22. Silvius, G. Harold, and Bohn, Ralph C. Planning and Organizing Instruction. Bloomington, Illinois: McKnight Publishing Company, 1976.

23. Smith, Robert G., Jr. The Engineering of Educational and Training Systems. Lexington, Massachusetts: Heath Lexington Books, 1971.

24. Smythe, Mary-Jeanette; Kobler, Robert J.; and Hutchings, Patricia W. "A Comparison of Norm-Referenced and Criterion-Referenced Measurement with Implications for Communication Instruction." The Speech Teacher 22 (1973).

25. Standards & Formats for Vocational Industrial Instructional Materials. College Station, Texas: Vocational Industrial Education, Texas A & M University, 1971.

26. Tanner, Daniel. Using Behavioral Objectives in the Classroom. New York: The Macmillan Company, 1972.

27. Thorndike, E. L. "The Nature, Purposes, and General Methods of Measurements of Educational Products." In The Measurement of Educational Products, edited by G. M. Whipple. Part II Seventeenth Yearbook of the National Society for the Study of Education. Bloomington Public School Company, 1918.

28. University of California, Division of Vocational Education. Foundations in Vocational Education: Reference and Work Book for Trade and Technical Teacher Education. Los Angeles: University of California, 1971.

Part II:

Content and Study Activities

PART II
CONTENT AND STUDY ACTIVITIES

Goal 9.1

Content Outline	Activities-Resources
Goal 9.1: Understand the Concept of Criterion-Referenced Measurement Within the Framework of Educational Evaluation.	
A. Criterion-Referenced Testing: Basic Definitions	(11) "A Criterion-Referenced Test," p. 41.
1. Various authors have defined the term "criterion-referenced testing." Let's examine some of those definitions now.	(15) Developing Vocational Instruction, p. 40.
2. According to Robert Glaser, "A criterion-referenced test is one that is deliberately constructed to yield measurements that are directly interpretable in terms of specified performance standards" (11).	
3. According to Mager and Beach, a criterion test "determines how well the student's performance at the end of instruction coincides with the performance called for in the objectives" (15).	
4. According to Kibler, Cegala, Barker, and Miles, a criterion-referenced test is "designed to determine whether a student has achieved mastery of a behavior as specified in an instructional objective" (14).	(14) Objectives for Instruction and Evaluation, p. 116.

5. According to Butler, the criterion test measures the individual's proficiency against a predetermined set of absolute criteria. Its main purpose is to determine as accurately as possible when a student has reached the acceptable level of performance (5).

(5) Instructional Systems Development for Vocational and Technical Training, p. 98.

6. According to Goldstein, "Criterion-referenced measures provide a standard of achievement for the individual as compared with specific behavioral objectives and therefore provide an indicant of the degree of competence attained by the trainee" (13).

(13) Training: Program Development and Evaluation, p. 63.

B. Measurement vs. Evaluation: Basic Definitions*

* See Discussion Question A in Part III.

1. It is necessary to distinguish between measurement and evaluation in order to put in perspective the concept of "criterion-referenced testing."

2. A criterion-referenced test is a "measuring instrument." "Measurement" refers to the activity of gathering and quantifying information through the use of a measuring instrument. No inferences, interpretations, judgments, or decisions are made about the information. The measuring instrument can be anything that collects raw data: teacher observation, a true-false test, a rating scale, an attitude scale, a personality inventory, an IQ test, or an anecdotal record kept by the teacher (7).

(7) Using Instructional Objectives in Teaching, p. 68.

3. Measurement is one activity in the more general process of evaluation. "Evaluation not only includes measurement but also the making of judgments and decisions based upon the gathered information. It is in evaluation, not measurement, that experience, judgment, and intuition enter the picture" (7).
4. According to Clark, "The majority of problems encountered by teachers as they evaluate can be directly linked to the quality and extent of their measurement. Inadequate evaluations are usually based upon faulty measurement or, in more extreme cases, upon little or no measurement. Though measurement is only one phase of the evaluation process, it is the basis from which the other phases stem" (7).
5. A criterion-referenced test, then, is a measuring instrument that measures learning outcomes in connection with instructional objectives.*

(7) Using Instructional Objectives in Teaching, p. 69. See also: (8) Home Economics Evaluation, Chap. 1.

* See Classroom Activity 1 in Part III.

C. Criterion-Referenced Measurement: An Historical Background

1. The concept of criterion-referenced measurement is not entirely new to educators. In 1918 Thorndike made reference to the distinction between the two types of measurement: norm-referenced and criterion-referenced. Thorndike, however, did not use these specific terms (27).

(27) "The Nature, Purposes, and General Methods of Measurements of Educational Products."

2. In 1963 Robert Glaser provided the initial conceptual clarity and indicated the practical implications of the two measurement procedures; his writing has stimulated numerous articles and papers elaborating on the applications, advantages, and liabilities of the two approaches (12).*

(12) "Instructional Technology and the Measurement of Learning Outcomes: Some Questions."

* See Discussion Question B in Part III.

Characteristics of Criterion-Referenced Tests and Norm-Referenced Tests

1. Various statements have been made in the literature comparing norm-referenced and criterion-referenced measurement. What follows is a summary of those statements.
2. Characteristics of Criterion-Referenced Measurement
   a. According to Smythe, Kibler, and Hutchings, "The main function of criterion-referenced measurement is to assess whether the student has mastered a specific criterion or performance standard.
   b. Complete instructional objectives are specified in the construction of criterion-referenced tests.
   c. The criterion for mastery must be stated (i.e., instructional objectives) for use in criterion-referenced measurement.
   d. Test items for criterion-referenced tests are constructed to measure a predetermined level of proficiency.
   e. Variability is irrelevant; it is not a necessary condition for a satisfactory criterion-referenced test.

f. The test results from criterion-referenced measurement suggest the use of a binary system (i.e., satisfactory/unsatisfactory; pass/fail). However, criterion-referenced test results can be transposed into the traditional grading system by following a set of specifically constructed rules" (24).

(24) "A Comparison of Norm-Referenced and Criterion-Referenced Measurement with Implications for Communication Instruction."

3. Characteristics of Norm-Referenced Measurement*

* See Classroom Activity 2 in Part III.

a. According to Smythe, Kibler, and Hutchings, "The main function of norm-referenced measurement is to ascertain the student's relative position within a normative group.

b. Either general conceptual outcomes, (usually done) or precise objectives may be specified when constructing norm-referenced tests.

c. The criterion for mastery is not usually specified when using norm-referenced tests.

d. Test items for norm-referenced measurement are constructed to discriminate among students.

e. Variability of scores is desirable as an aid to meaningful interpretation.

f. The test results from norm-referenced tests are amenable to transposition to the traditional grading system (A, B, C, D, F)" (24).

(24) "A Comparison of Norm-Referenced and Criterion-Referenced Measurement with Implications for Communication Instruction."

4. Desirable Characteristics of a Criterion-Referenced Measuring Instrument

a. A good criterion-referenced test will be valid, reliable, objective, comprehensive, and economical. (Although instructors

and curriculum specialists are most familiar with these terms in regard to norm-referenced tests, such characteristics are desirable for criterion-referenced tests as well. However, these terms take on somewhat different and special meanings when used in regard to criterion-referenced tests.)

b. Validity. If a criterion-referenced measuring instrument requires the same behaviors that are identified in the objectives, then the scores are said to be valid. (Less precisely, the instrument is said to be valid.) According to Clark, "An objective asks the student to demonstrate some behavior relative to some content; a measuring instrument also asks the student to demonstrate some behavior relative to some content. The degree to which the two behaviors and the two topics correspond will be the degree to which the instrument is valid. This type of validity is called 'content' validity (7). Clark concludes that of all the desirable characteristics of a measuring instrument, content validity is the most important. If a measuring instrument generally fails to measure what it was designed to measure, all other characteristics lose their meaningfulness. To assure content validity in their measuring instruments, course developers and instructors should

(7) Using Instructional Objectives in Teaching, p. 69.

make certain their objectives are clearly defined, for the objectives provide the standards for making judgments about the validity of the measuring instruments.

c. Reliability. A criterion-referenced test that is reliable will be consistent in its measurement. It will measure in exactly the same way every time it is used. For example, if a test is given to a specific group of students one day and then given again to the same group on another day, the test scores should be relatively the same for the individuals on both days.

If the objectives call for behavior that is observable and measurable and the test items call for this same behavior, the test will probably have a high degree of reliability (5)

(5) Instructional Systems Development for Vocational and Technical Training, p. 100.

d. Objectivity. A good criterion-referenced test must be relatively objective, that is, the judgment of the scorer should enter the scoring process as little as possible. The scores on a good test will be about the same regardless of the individual doing the scoring. All else being equal, an objectively scored test (which does not permit scorer bias to affect the score) is more valid and reliable than a subjectively scored one (5).

e. Comprehensiveness. With a test item for every objective, the criterion-referenced test will necessarily give comprehensive

coverage of all desired behaviors.

f. Economy. The criterion test must be economical regarding time, manpower, and facilities; but economy is strictly relative when applied to an instructional system. The gain in reliability and validity through the use of a truly comprehensive, criterion-referenced performance test far outweighs the economy of group paper-and-pencil tests (5).*

(5) Instructional Systems Development for Vocational and Technical Training, p. 100.

* See Discussion Question C in Part III.

E. Standardized Tests

1. The majority of schools in the United States today make use of standardized tests of one kind or another. Most tests of intelligence, aptitude, personality, and interests are standardized tests, made by specialists for a test publisher, and sold by the publisher throughout the country. Few schools or school systems, except in very large city organizations, attempt to develop such tests for their own use.

2. The situation with respect to achievement tests is somewhat different. There are, of course, many standardized achievement tests on the market, and literally millions of them are used every year. These include tests in the separate subjects or branches in addition to the achievement batteries. However, teachers usually feel that these tests do not adequately measure their own or the local objectives of instruction. Thus, while standardized tests are very useful in some ways, they are not usually the principal method of measuring

achievement. In general, it is the classroom teacher or curriculum specialist who is relied upon to formulate achievement tests. It is important, therefore, that the teacher's and specialist's professional training include some instruction on effective ways of planning, constructing, and evaluating various measuring instruments.

3. Clearly, no standardized test of achievement can serve the needs and purposes of every local situation. The requirements for a standardized test are such that the test must be largely confined to instructional elements common to a large number of schools. Such a test cannot, therefore--if it is to be maximally useful--include all those elements that are peculiar to any one or even to a limited number of schools. The most desirable and probably the most common practice is to use both standardized and teacher-made measuring instruments in most situations (28).
4. Today, criterion-referenced achievement tests are being "standardized," that is, developed by specialists for a test publisher and sold throughout the country to a large market. Such tests will have the same problems as other standardized achievement test They won't adequately test the unique objectives of the local situation. Teachers and curriculum specialists will continue to rely largely on the development of their own tests. When

(28) Foundations in Vocational Education: Reference and Work Book for Trade and Technical Teacher Education.

the occasion does arise to select a standardized criterion-referenced test, however, the teacher or curriculum specialist must be cautious that the behavioral criteria of test items are spelled out clearly, that in fact the test items measure what they say they are intended to measure.*

* See Classroom Activity 3 in Part III.

F. Study Activities

Based on your reading of the content outline and any additional references as suggested, complete the following activities.

Basic Definitions (19)

Educational evaluation: Educational evaluation refers to the determination of the worth of educational phenomena; it generally refers to the evaluation of an educational enterprise, such as an instructional sequence, not to the evaluation of students within that enterprise. Educational evaluation is a process of worth determination.

Educational measurement: Educational measurement refers to the assessment of the current status of an educational phenomenon in a precise fashion--that is, counting or enumerating so that the phenomenon can be more accurately described--without placing value (goodness or badness) on the phenomenon thus described. Educational measurement is a process of status determination.

Criterion-referenced testing: Criterion-referenced testing is a form of educational measurement that ascertains an individual's status with respect to some criterion or performance standard. Because the individual is compared with some established criterion, rather than with other individuals, these measures are described as criterion-referenced.

Norm-referenced testing: Norm-referenced testing is a form of educational measurement that ascertains an individual's performance in relationship to the performance of other individuals on the same measuring device. Because the individual is compared with some normative group, such measures are described as norm-referenced.

1. Determine whether the following examples represent educational evaluation or educational measurement by marking an "X" at the appropriate choice.

a. A vocational counselor directs a testing program that provides IQ scores and comprehensive achievement scores for each student in the district.

_____a. evaluation

_____b. measurement

b. A vice-principal in an area vocational school observes a home economics class for a week and concludes that the instructor lectures too much, providing little opportunity for students to participate.

_____a. evaluation

_____b. measurement

c. An industrial arts instructor administers an examination to determine if the students in his class have achieved the instructional objectives for a unit of instruction.

_____a. evaluation

_____b. measurement

d. The principal of a comprehensive high school annually determines the comparative percentile ranks of all entering freshmen in English and mathematics.

_____a. evaluation

_____b. measurement

2. Determine whether the following examples represent criterion-referenced testing or norm-referenced testing by marking an "X" at the appropriate choice.

a. In the Red Cross Senior Lifesaving Test, an individual must demonstrate certain swimming skills to pass the examination, regardless of how well others perform on the test.

_____a. criterion-referenced testing

_____b. norm-referenced testing

b. Although a business student scored 90% on an examination, he did not receive an A because a majority of the students in the class scored higher.

_____a. criterion-referenced testing

_____b. norm-referenced testing

c. A test is used to determine the top 25 vocational students for a new vocational leadership program.

_____a. criterion-referenced testing

_____b. norm-referenced testing

d. Students in a woodworking shop are required to pass a knowledge test of basic safety rules before operating any equipment in the shop.

_____a. criterion-referenced testing

_____b. norm-referenced testing

3. Read the "Foreword" and Robert Glaser's article "Instructional Technology and the Measurement of Learning Outcomes: Some Questions" in Popham, Criterion-Referenced Measurement: An Introduction. Then complete the following questions.

   a. How did World War I psychology promote the use of norm-referenced measurement in education?

   b. Who coined the term "criterion-referenced measurement" and when?

   c. What factors in education do you think contributed to an increasing emphasis on criterion-referenced measurement?

   d. What form of measurement is primarily used in education today? How would you explain this phenomenon?

   e. What is Glaser's primary concern regarding the measurement of learning outcomes?

4. Read the Special Report of the Association of California School Administrators on "The Nature and Uses of Criterion-Referenced and Norm-Referenced Achievement Tests," provided on the following pages. Then complete the questions below.

   a. In what sense is the concept of criterion-referenced measurement "new"?

   b. What are other current terms in use for criterion-referenced tests?

   c. Under what circumstances might a criterion-referenced test be considered "standardized"?

SPECIAL REPORT

Vol. 4, No. 3

Research & Evaluation Committee
Robert Hocker, Chairperson
Subcommittee:
Donald Ross Green, CTB/McGraw-Hill
Norman Ginsburg, Ocean View SD
Harold Hyman, Compton Unified SD

THE NATURE AND USES OF CRITERION-REFERENCED AND NORM-REFERENCED ACHIEVEMENT TESTS

Introduction

During the past few years, the term "criterion-referenced test" has come into increasing wide use to the point where it is virtually a buzz word. Almost any discussion among school administrators of "what's new" will include allusions to criterion-referenced tests and perhaps domain-referenced tests, mastery tests, or objectives-based tests as well. These sorts of tests are usually contrasted with norm-referenced tests, standardized tests, traditional tests, or the like. It is particularly common to find that such discussions are confusing because people have different ideas about what these words and phrases mean and about what the various tests are good for.

Most school officials hear criticism of their testing programs from all sides and consequently many are finding the various claims about criterion-referenced tests both enticing and disturbing. Are they the answer to the teachers' prayers? The principals'? What are they really? Just another fad? Why all the different terms? Can we now discard the traditional achievement tests? Are there handy books or articles answering these questions for school administrators? The answer to the last of these questions is "no" and hence this paper.

Thus, the general purpose of this report is to enable general administrators to discuss these matters more knowledgeably with their staff and communities. The specific purposes are to (1) clarify this proliferating terminology (and perhaps stunt its growth) by enumerating common differences and distinctions in meaning between terms and between usages of these terms, (2) explain the essential nature of the difference between the two kinds of tests, and (3) offer a viewpoint on their uses in schools. The discussion is limited to achievement tests, and emphasizes what the terms imply about differences in how tests are constructed and how they may be used. It is hoped that readers of this report will be better able to infer what is meant when they encounter discussions of these ideas and will be better able to judge and use the various kinds of measures available. It may even be hoped that the result will be a general reduction in confusion.

The confusion arises because there is a diversity of usage among the "experts." In this context, an expert is anyone who has written something about the two contrasting kinds of tests in a recognized journal or in a book. The contrasts and distinctions in usages and meanings cited in this report are largely composites and cannot be attributed to any one source.*

*A list of the major references consulted is appended. Furthermore, we wrote to a number of scholars and organizations asking for their definitions. Those replying were: Professors J. Stanley Ahmann, NALP; Robert Ebel, Michigan State University; William Merz, Sacramento State College; Jason Millman, Cornell University; M. I. Chas. E. Woodson, University of California, Berkeley; and the Center for the Study of Evaluation, UCLA; CTB/McGraw-Hill; and Houghton Mifflin. We are grateful for these replies; all of them were thoughtful and useful. We might add, they make us feel confident that the problem just outlined does indeed exist.

SPECIAL REPORT is published by the ASSOCIATION OF CALIFORNIA SCHOOL ADMINISTRATORS

"What's New?"

Neither norm-referenced nor criterion-referenced tests are particularly new. The Chinese used a species of norm-referenced tests for hundreds of years in their civil service program, and teachers in many places have used various kinds of criterion-referenced tests in their classes. Both kinds of tests have undergone changes and technical development in the last century. It is in the technical sense that criterion-referenced tests are rather new. Only in the last few years have those concerned with theoretical and technical issues of educational and psychological measurement, i.e., psychometricians, undertaken any sustained large scale effort to deal with criterion-referenced tests. It is even more recent that publishers of tests for schools have tried to offer any substantial tests of this sort. Therefore, consensus among the experts about technical requirements for the tests has yet to be reached and little about these matters can be found in courses and textbooks.

To explore these issues, we first offer crude definitions of criterion- and norm-referenced tests. Then we try to dispose of some "red-herring" problems created by the words themselves and consider the alternative terminology that has been suggested. Then the basic differences between the two kinds of tests are pointed out and related to some important technical issues in construction and test interpretation. The final section of the paper pertains to appropriate and effective uses of these two species of achievement tests.

Definitions

Norm-Referenced Measures

Most writers agree that norm-referenced achievement tests use a sample of questions that refer to a broadly defined set of educational goals. Scores are meant to tell how much the student knows about that area or what level of ability he or she has attained. Because of the lack of a fully defined (enumerated) body of knowledge (try to list all knowledge about American history) or of a natural bottom and top level of ability (what is 100% computational skill?), score meaning is most readily obtained by comparing scores of students. If a well-defined group of students (the normative population) is sampled properly, any score can be compared with those of the rest of the population. The data derived from the scores of the sample, which permit these comparisons, are the norms. They may appear in many forms such as standard scores, percentiles, or grade equivalents. In any case, the norms are a consequence of the basic nature of the test; they do not determine that nature.

Criterion-Referenced Measures

A criterion-referenced achievement test provides a set of questions that refer to relatively restricted (i.e., specific) educational objectives. Performance on the items is meant to tell how much the student knows about the topic or how well he or she can perform the task. The specificity of the objective and the clarity with which it describes the behavior representing the objective to be achieved provides direct meaning to the scores. For example, the student knows (or does not know) how to identify the topic sentence of a paragraph. Criterion-referenced tests usually measure a large number of objectives, each treated separately. Since objectives are not equally important and some are contained in others, simply adding the scores together will not yield a score that means what the score on a regular achievement test does.

Problem Words

Norms

Creation of norms for a criterion-referenced test is entirely possible. Just because the test is built to be directly interpreted relative to some performance standard does not preclude its also having norms and being used as a norm-referenced test. Especially if it is to be used for program evaluation this may be a reasonable procedure. However, since the test has probably been designed to be used for instructional planning and guidance, it is less likely to be as useful for other purposes. In any case, the existence of norms does not make the test a norm-referenced test. Fortunately, this source of confusion is rare.

Criterion

Analogously, one may define some score on a test constructed to be norm-referenced as the criterion score indicating mastery. Since the breadth of the scale usually makes this decision arbitrary and not something others would automatically

understand, it merely creates confusion to say that this makes a test criterion-referenced. This sort of confusion is widespread and is complicated by the difficulty entailed in specifying the breadth of an objective or its domain. The comments of those who take this position indicate that they really maintain that all tests are alike and ultimately must be norm-referenced. This position is contradicted by the common sense interpretation used by many teachers with their own classroom tests.

Standardized

There is little unanimity of opinion about the meaning of the word standardized as a descriptor of tests. For some, it merely means tests with norms. For others, it means the test is (1) published, (2) normed, (3) has explicit instructions for administration, and (4) was constructed to meet technical standards. Still others leave out requirement 1 or requirement 2 or both. By this last definition, many criterion-referenced tests are standardized. Plainly, this term is not useful unless one specifies the definition being used whenever the word is employed. The most common useful definition (and the one that we use) addresses itself to test administration and scoring. Strict adherence to the author's or publisher's administration and scoring directions is necessary if reliance is to be placed on the results. Failure to adhere to them means that norms cannot be used, comparisons with previous testings or other groups cannot be made, and so on.

Alternatives to the Term "Criterion-Referenced"

Let us return to the definitions of norm-referenced and criterion-referenced tests. It is apparent that the ideas are clear enough but the terminology is unfortunate. This is why all the other labels for criterion-referenced achievement tests exist. Yet it seems clear to us that matters have gone too far to change; the terms "criterion-referenced" and "norm-referenced" are too well established to be abandoned. The writers who use other words have excellent reasons for doing so, but usage is against them and they do not agree with each other. They do agree there are two general kinds of achievement tests although some of those responding to our questions preferred to describe these differences as a matter of degree and emphasis rather than kind. Nevertheless, it seems to us that a general consensus exists about the underlying ideas but that differences in terminology tend to obscure these agreements.

Those who prefer the term "domain-referenced" are concerned about test construction procedures (see the next section) while those who use "mastery" emphasize a use. The term "objectives-based" elicits wide agreement but little enthusiasm. Generally speaking, those who use terms other than criterion-referenced appear to understand perfectly well what is meant, they simply do not like the terminology. Therefore, if one accepts the proposition that usage has already established norm-referenced and criterion-referenced, the problem is not those few who would use other labels, but the many who have erroneous notions about what these two terms mean. Let us, therefore, try to elaborate on the nature of these two kinds of tests.

Characteristic Differences Between Criterion- and Norm-Referenced Tests

The two kinds of tests may both have norms, may both have a particular criterion score designated as indicative of mastery, and may both be considered standardized. The two kinds often differ on these points but they do not have to do so. The basic differences arise in the construction procedures.

First there are the content specifications for test construction from which content validity is established. The ideal process of building a traditional standardized achievement test is well known. Many textbooks describe it. The first step is to establish content specifications; experts in the area outline the topics to be included and indicate their relative importance. Items are written to fit the topics; importance of a topic is reflected by the number of items about the topic that are included. The test is usually designed to yield a few (e.g., 4 to 7) separate sub-test scores each based on perhaps 20 to 50 items, as well as a total score.

For criterion-referenced tests, the content specifications procedure is similar except that each objective or component of the topic to be measured is considered separately. The specifications indicate all the topics but give them no weights. A criterion-referenced reading test, for example, may include 40 to 60 objectives each of which is measured by, perhaps five items. In effect, there are 40 to 60 separate tests and scores but no total score.

Content validity is judged in both cases by examining the adequacy of content coverage, but in the traditional test the appropriateness of the emphases is an important criterion for judging. Content validity of both types of tests is also judged by examining how well the items fit the categories or objectives they are meant to measure.

In short, the only difference in content criteria is in the weighting and the breadth of content represented by any one score. This difference may not appear large, but it represents two very different views of the measurement task. For norm-referenced tests, there is assumed to be a trait or capability (e.g., reading skill) that individuals have in different amounts because of different amounts of learning. The measurement task is to place people on a scale of that trait; the scale is usually established by norms. For criterion-referenced tests, a behavior is described that occurs under certain conditions when a student has achieved the objective; the measurement task is to determine if a student has achieved the objective. When objectives are complex and broad, one can measure degrees of achievement and the two sorts of tests begin to look alike. However, as long as it

is possible to interpret scores as direct descriptions of achievement, the test can be called "criterion-referenced."

The second step in test construction is to write items to fit the content specifications. Generally speaking, the process does not differ for the two sorts of tests. However, some writers feel strongly that they should differ. These are the advocates of the term "domain-referenced," which refers to their item writing procedures.

The third step in test construction is to try out the items and select the best ones to make up the test. Here the differences between the two kinds of tests are more numerous. On standardized achievement tests, the "best" items are usually those that: (1) discriminate well, i.e., are answered correctly by generally high scoring students and incorrectly by low scores (the point-biserial coefficient — an item-test correlation — is one common index of this); (2) show growth from grade to grade (most achievement batteries provide different levels of the test for every one, two, or three grades); (3) are about in the middle in difficulty — e.g., on a multiple-choice test, about 65% of an average group would get the item right (this is the average one aims for but it is helpful to have some easy items and some hard ones); (4) if the item is multiple-choice, each alternative (distractor or foil) should be chosen by many of the lower scoring students.

In criterion-referenced tests, by contrast, one looks for an item that will discriminate between students who have and have not achieved the particular objective it is measuring but there is no concern with its relation to other objectives. One also looks for items that will show mastery immediately after the objective is achieved. Thus, if an objective is taught during a week, many if not most students should fail the item before that week's instruction and pass it after (assuming that the instruction is adequate). This characteristic is called "sensitivity to instruction." Determination of sensitivity to instruction requires a two-stage tryout, one before and one after relevant instruction; a single-stage tryout with a treatment (taught) and a control (not taught) group can provide similar information provided that assignment to treatments is done properly. A traditional achievement test has only one tryout and its timing relative to instruction is rarely considered. Selection of items that are sensitive to instruction means that a criterion-referenced test will reflect learning as it happens, something that regular achievement tests rarely do. It takes six or eight months of schooling for most norm-referenced test scores to reflect significant changes because they measure many broad and complex goals all at once. This is one reason why criterion-referenced tests can be more effective than norm-referenced tests when used to guide instruction and for looking at programs internally while in progress. Norm-referenced tests on the other hand are efficient when used to evaluate program outcome because of their long-term summative-character. Finally, one looks for items that will be difficult for those who have not achieved the objective and easy for those who have. "Middle" difficulty items are not particularly desirable. This is equivalent to the preceding point but contrasts with norm-referenced items.

Table 1 summarizes some of the similarities and differences between the two kinds of achievement tests. Note that most of these characteristics are more properly labeled "typical" than necessary. These differences in nature and construction point to different uses of the two kinds of tests.

Table 1
Characteristic Similarities and Differences between Norm-Referenced and Criterion-Referenced Achievement Tests

Norm-Referenced Tests	Criterion-Referenced Tests
Content Specifications	
1. Topics outlined and weighted according to importance; number of items per topic is directly proportional to importance.	1. Topics broken down into specific educational objectives; number of items per objective is usually constant. In any case, all objectives are equally represented since each has its own score.
2. Both omission of important content and inclusion of unimportant content are serious flaws that distort meaning of scores.	2. Omission of important topics reduces overall value of instrument but does not affect meaning of scores. Unimportant objectives can be ignored.
3. Test usually covers broadly defined educational goals that represent the most widely adopted school curricula.	3. Test covers a set of specific educational objectives.
4. Altering a test to fit a specific local curriculum is very difficult; it is usually easier to build such a test from scratch.	4. The set of objectives used may be easily selected or modified to fit local curricula.
Item Writing Specifications	
1. Items are usually written to learning objectives which represent a sample of those relevant to the goals being measured. Each goal is systematically sampled but objectives are not.	1. Items are written to learning objectives; each objective is systematically sampled.
	2. Items refer only to the objective to which they written

2. Single items often require knowledge of several aspects of the content.

Desirable Item Characteristics

The best items are those that:	The best items are those that:
1. discriminate well between those who score high and those who score low on the test,	1. discriminate between those who have and have not had effective instruction to that objective,
2. show growth from grade to grade,	2. show mastery immediately after the objective has been achieved,
3. are about midrange in difficulty (but some items at each extreme are also desirable).	3. have preinstruction difficulties approaching 0 (almost all get them wrong) and postinstruction difficulties approaching 1 (almost all get them right).

Administration

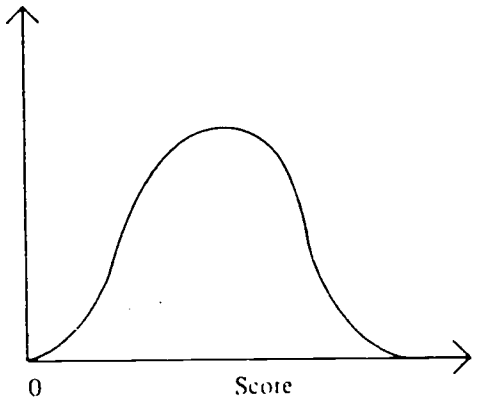
1. Standardized conditions of administration are essential including control of time (sometimes tests are speeded but not always).	1. More latitude in conditions is permissible. Control of time is rarely appropriate (unless speed is part of task).
2. Parts cannot be omitted without damage to meaning of total.	2. Parts can be omitted at will since there is no total score.

Scores

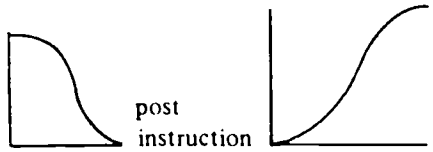
1. Raw scores rarely have much direct meaning.	1. Raw scores have some direct meaning about achievement of the objective being measured.
2. Measurement places person on hypothetical scale of amount of trait.	2. Measurement refers to scales based on visible performance.
3. Scale usually established by norms (comparative performances).	3. Scale is usually established by judgment and convention concerning adequate and inadequate performance but norms may exist and help.
4. Derived scores are used such as standard scores, percentile ranks, grade equivalent scores.	4. Scores used are number right, and categories such as mastery and nonmastery.
5. Score reports usually imply value, i.e., performance was good or poor.	5. Score reports are less well adapted to making conclusions about the quality of student or program performance.
6. All items contribute to part and total scores.	6. Each objective has its own score; meaningful total scores are usually not possible.

Score Distributions

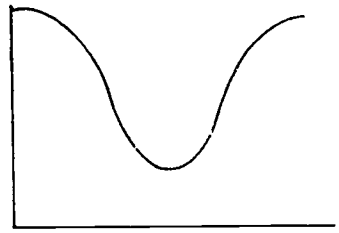
1. Score distributions that are approximately normal are desirable, e.g.:



1. Score distributions that are skewed are desirable, e.g.:



2. If the group tested includes both preinstructed and uninstructed students, distribution should be U-shaped, e.g.:



Reliability

1. Test-retest coefficients should be high for each score.

1. Test-retest coefficients should be high for each objective in a mixed sample (as above).

2. Internal consistency coefficients should be substantial for each score.	2. Internal consistency coefficients should be high for each objective in a mixed sample.
Content Validity	
1. Content coverage and emphasis should be judged adequate.	1. Adequacy of coverage of behavior specified by objective should be adequate.
2. Fit of items to their intended content category is a matter of judgment.	2. Fit of items to their intended content category is a matter of judgment.
Construct Validity	
1. Scores show growth during years of school attendance.	1. Scores for objectives exhibit sensitivity to instruction, i.e., change from wrong to right after effective instruction.
2. Scores show greatest growth during years of relevant instruction.	2. Items for one objective are more closely related than across objectives.
3. Groups with more training average better than groups with less.	3. General background plays less role than in norm-referenced tests (this implies less cultural bias).
4. High scoring students can more often solve problems requiring the knowledge than low scoring students	4. High scoring students can more often solve problems requiring the knowledge than low scoring students.
5. Relationships among items should correspond (show patterns) to relationships among content categories (e.g., results of factor analyses should be logical).	5. Relationships among items should correspond (show patterns) to relationships among content categories (e.g., results of factor analyses should be logical).
	6. Where a learning hierarchy is known to exist performance on higher objectives will predict performance on lower order objectives, and demonstrated mastery of lower order objectives facilitates learning of higher order objectives (e.g., positive vertical transfer).
Criterion Related Validity	
1. Scores correlate well with other measures of achievement such as teachers' marks and other tests.	1. Scores correlate well with other measures of the objective.
2. Scores predict performance in class or on tasks dependent on capabilities being measured.	
Uses	
1. Assessment of status of school system (or classes or students) with respect to achievement in basic skills and content areas.	1. Assessment of status of students (or classes or school system) with respect to curriculum objectives.
2. Program evaluation for outcomes of long-term growth (at least 6 months) towards major goals.	2. Program evaluation for long- or short-term attainment of specific objectives.
3. Selection and placement of students in courses and programs on the basis of level of basic skills or general knowledge of content.	3. Diagnosis of instructional needs of individual students and groups of students.
4. Information for curriculum planning.	4. Information for planning of classroom instruction.
5. Monitoring yearly progress of schools and school systems with respect to goals.	5. Monitoring progress of students with respect to instructional objectives.

Appropriate and Effective Uses

There is a wide variety of ways one may design testing programs using norm-referenced and criterion-referenced tests. Because the possibilities are so many, we shall describe a particular approach now being tried in the Ocean View School District of Orange County. We believe it is a conceptually sound approach and is feasible in at least some systems. Other approaches equally effective unquestionably exist. Although our purpose is to illustrate possibilities, not prescribe, we do have some convictions about test programs which this approach reflects.

In particular, we believe that in this age of individualized instructional programs school districts throughout California should be using both kinds of tests in a complementary sense. This is because of the large variety and quantity of information that is needed if such programs are going to function adequately. On the one hand school administrators need to know how well their students are performing collectively in the basic skill areas vis à vis a benchmark independent of their district. Without this information it is not possible for them to report responsibly to the school board, to parents, or to the public. By the same token without such information these latter groups cannot make responsible decisions about the policies and support they must provide the schools. Without this information program planning, or revision and improvement is handicapped.

On the other hand, a different sort of information is needed for planning and evaluating instructional activities. Site administrators and program supervisors but especially teachers and students need specific information about student achievement of instructional objectives. Bright students, sometimes helped by well educated parents, often figure out for themselves just where they stand but many students need unambiguous feedback from the teacher about their efforts to learn. Teachers who try seriously to do this on some real basis other than rule of thumb need help in getting this information especially if they are trying to individualize instruction.

We believe that tests are the most reliable, valid and efficient source of much (although not all) of these kinds of information. Carefully constructed and validated tests properly used can provide information good enough to say that the time and money required to administer the tests is worthwhile and that the program or instructional decisions are more likely to be correct than if the decision maker chose not to test. Thus district administrators, site administrators, and teachers all need to know how to use tests effectively. They need to know how to proceed efficiently and how not to interfere with each other's efforts or with student learning. The Ocean View program is intended to achieve this condition.

The Ocean View Program

Ocean View School District has had a well established commitment to tests and program evaluation for a number of years. From 1969 to 1974, despite shrinking budget options, the district has expended between $3 and $5 per student per year on tests and related evaluation activities to meet the consistent and insistent demands for cognitive student information. The school district has consistently exceeded state requirements for testing for the benefit of its students, teachers, administrators, parents, school board, and taxpayers.

In Ocean View as elsewhere the school board has the responsibility to identify the general parameters of the curriculum. Working from these general guidelines, curriculum directions and strategies must be identified at the administrative and teaching level. Concomitantly, evaluation instruments and strategies are identified to fit the curriculum parameters adopted by the board. Like most school districts, Ocean View uses a norm-referenced test to gain a broad picture of how well the district is achieving in its curriculum in the basic skills areas of reading, language, and mathematics. In those areas in which the district is doing well, it is assumed that district teaching efforts are effectively meeting the goals of the educational program. If, on the other hand, the district shows a weakness, an in-depth study of the problem is then initiated. A norm-referenced item analysis can be part of such a study, however, when one needs to translate identified group program weakness into individual student profiles for corrective action, a criterion-referenced test is the best measure.

Therefore, the district program includes criterion-referenced tests. Criterion-referenced tests with their interim tests* can assist teachers, regardless of grade level, to know on a regular basis how well a student is doing in relation to subject area objectives. Criterion-referenced tests indicate specific student educational deficits or strengths on which the teacher can base a prescription of daily educational tasks for that student. Individual instructional programs can be derived from criterion-referenced test data if the objectives which the test measures are explicitly those in the curriculum plan. In this sense teachers teach to the test.

In short, Ocean View has concluded that a basic test program should include both norm-referenced and criterion-referenced tests. The district uses one of the traditional norm-referenced achievement batteries in Grades 1 through 8 to obtain a broad picture for its program evaluation. The intention is to use criterion-referenced tests for diagnostic/prescriptive purposes with individual students. In the area of mathematics, a published criterion-referenced test has been adopted and is now in use. The test results are used to translate general norm-referenced information into a meaningful instructional picture for both the teacher and site administrator. In reading, the district criterion measurement committee is currently exploring various published programs that can be adapted to meet Ocean View's needs.

*Interim tests are short quizzes designed to measure a single objective and are usually parallel to their corresponding part in a more comprehensive criterion referenced test. They are useful for monitoring the daily or weekly progress of individual students.

Difficulties in Effective Usage

There are problems generated by an extensive testing program. One is the recurring professional argument as to whether a norm-referenced achievement test should regulate the curriculum. Most say it should not. Nevertheless, if one is attempting to measure achievement, the test content must have a close relationship to the curriculum taught. In fact, the Ocean View experience suggests that it is not too hard to get a substantial degree of consensus from teachers concerning the appropriateness of most of the items found in the basic skills measures used. When teachers question whether a given achievement test is regulating the curriculum, it is helpful to ask them to examine the particular items and the skills they are intended to measure. Specific disagreements about relevance and appropriateness are rarely found. Skills such as knowing how to identify main ideas, how to use punctuation properly or how to perform basic arithmetic operations are pretty nearly universally accepted. The broad general content of a norm-referenced test is so common in agreement vis à vis appropriate curriculum areas in public schools that getting agreement is not difficult when the time is taken to do so.

Using both kinds of tests creates a second problem, namely interrelating the results of the two kinds of instruments. The two do yield somewhat different kinds of information: an overall summative judgment on the one hand and a specific diagnostic picture on the other. It is important to help personnel see the interrelationships among the two sets of data and not come to feel they are in conflict. This also calls for careful examination of test materials by teachers and administrators. By carefully determining correspondences and differences in content and emphasis, criterion-referenced tests can be used to translate general norm-referenced information into a meaningful instructional picture. The strengths and weaknesses of both kinds of information can be exhibited when this is done. It is evident that administrators must help teachers in this task.

The third problem occurs in the subsequent task of relating the information to instructional activities. The quantity of information is large and teachers cannot use it without help. A particular need arises to assist teachers in a supportive sense as they gather materials and equipment and organize their students for instruction.

Finally we may add that when a large degree of usefulness of the district testing program is developed the problems created by external testing demands are reduced. This is because students and teachers (a) perceive tests as having some use and value and (b) tend to understand more fully what the data do and do not mean with a consequent reduction in fear and distaste.

Conclusion

In brief, the key to the effective use of tests is a balanced use of each type of test. Norm-referenced tests are used best for group decisions. Criterion-referenced tests are used best for individual diagnostic/prescriptive decisions.

When evaluation questions arise concerning comparisons of groups of students for purposes of making general curriculum decisions, norm-referenced tests are usually most appropriate. When evaluation questions arise concerning the educational progress of a student for purposes of prescribing an individualized instructional program, criterion-referenced tests are usually most appropriate.

REFERENCES

Carver, R. P. Two dimensions of tests: Psychometric and edumetric. American Psychologist, 1974, 29, 512-518.

Ebel, R. L. Criterion-referenced measurement: Limitations. School Review, 1971, 79, 282-288.

Glaser, R. & Nitko, A. J. Measurement in learning and instruction. In R. L. Thorndike (Ed.), Educational Measurement. Washington, D. C.: American Council on Education, 1971.

Hambleton, R. K., & Novick, M. R. Toward an integration of theory and method for criterion-referenced tests. Journal of Educational Measurement, 1973, 10, 159-170.

Harris, C. W., Alkin, M. C., & Popham, W. J. (Eds.) Problems in criterion-referenced measurement. CSE Monograph Series in Evaluation, Number 3. Center for the Study of Evaluation, Univ. of California, Los Angeles, 1974.

Hively, W., Maxwell, G., Rabehl, G., Sension, D., & Lundin, S. Domain referenced curriculum evaluation: A technical handbook and a case study from the MINNEMAST project. CSE Monograph Series in Evaluation, Number 1. Center for the Study of Evaluation, University of California, Los Angeles, 1973.

Klein, S. P., & Kosecoff, J. B. Issues and procedures in the development of criterion-referenced tests. ERIC/TM Report 26. Princeton, NJ: ERIC Clearinghouse on Tests, Measurement, & Evaluation, 1973.

Kosecoff, J. B., & Klein, S. P. Instructional sensitivity statistics appropriate for objectives-based test items. CSE Report, No. 91. Center for the Study of Evaluation, University of California, Los Angeles, 1974.

Popham, W. J. (Ed.) Criterion-referenced Measurement: An Introduction. Englewood Cliffs, NJ: Educational Technology Publications, Inc., 1971.

Roudabush, G. E. Item selection for criterion-referenced tests. Paper presented at the annual meeting of the American Educational Research Association, New Orleans, February, 1973.

Smith, C. W. Criterion-referenced assessment in contrast to norm-referenced measurement. Marquette University Education Review, 1973, 1, 25-33.

5. Having read the Special Report of the Association of California School Administrators on "The Nature and Uses of Criterion-Referenced and Norm-Referenced Achievement Tests," complete the following exercise by deciding whether the characteristics indicated describe a norm-referenced test or a criterion-referenced test. Mark an "X" at the appropriate choice.

a. With this type of test, the test items refer only to the specific objectives for which they were written.

_____a. norm-referenced test

_____b. criterion-referenced test

b. With this type of test, the best test items are those that discriminate between individuals who score high and individuals who score low on the test.

_____a. norm-referenced test

_____b. criterion-referenced test

c. With this type of test, the number of test items per topic is determined on the basis of the relative importance of that topic.

_____a. norm-referenced test

_____b. criterion-referenced test

d. This type of test determines whether or not students have achieved specific instructional objectives.

_____a. norm-referenced test

_____b. criterion-referenced test

e. This type of test determines a student's status with respect to other students in the achievement of basic skills and content areas.

_____a. norm-referenced test

_____b. criterion-referenced test

f. With this type of test, meaningful total scores on all test items are usually not possible.

_____a. norm-referenced test

_____b. criterion-referenced test

6. Read the Popham and Husek article, "Implications of Criterion-Referenced Measurement" in Popham, Criterion-Referenced Measurement: An Introduction. Then complete the following exercise by distinguishing between norm-referenced measurement (NRM) and criterion-referenced measurement (CRM) on the bases indicated below.

   a. VARIABILITY

      CRM and NRM

   b. ITEM CONSTRUCTION

      CRM and NRM

   c. RELIABILITY

      CRM and NRM

   d. VALIDITY

      CRM and NRM

   e. ITEM ANALYSIS

      CRM and NRM

   f. REPORTING AND INTERPRETATION

      CRM and NRM

(See Appendix A for possible answers.)

Content Outline	Activities-Resources

Goal 9.2: Select Approaches/Techniques For Assessing Student Achievement of Instructional Objectives in the Three Domains of Learning.

A. Measuring Instruments for the Cognitive Domain

1. Written tests, whether they be teacher-made or standardized, play a central role in the testing of knowledge.
2. According to Butler, there are three basic types of criterion test items that can be used to test knowledge. These types are:
   a. "A test item may be directive or imperative. For example, 'Find the value of R in an electrical circuit if I=30 amperes and E=110 volts.'
   b. An item may be a completion type, requiring the student to select from several possible choices the one that he thinks will correctly complete the stem of the item. For example, 'Excessive backlash in the differential assembly of an automobile would most likely be caused by'... (followed by a blank space or four alternative responses).
   c. An item may ask a direct question. For example, 'What three types of meter functions are combined in a multimeter?'" (5).

(5) Instructional Systems Development for Vocational and Technical Training, p. 195.

3. These three basic types of criterion test items can be adapted to measure all the different kinds of acquired knowledge. As Butler points out, however, it is important to measure the students' knowledge by testing their ability to apply that knowledge to the problems they will encounter on the job, rather than by the mere recall of isolated facts. To do this, Butler strongly recommends the method of providing hypothetical situations and then asking practical, objective questions about courses of action that should be taken in the situations (5).

(5) Instructional Systems Development for Vocational and Technical Training, p. 195.

B. Measuring Instruments for the Affective Domain

1. According to Pucel and Knaak, attitude measurement is one of the more complex types of measurement. People have attempted to measure attitudes for years and have developed very complex assessment procedures that have had only minimal success (21).

(21) Individualizing Vocational and Technical Instruction, p. 186.

2. According to Armstrong et al., "Perhaps the most appropriate instruments for measuring development toward course objectives in the affective domain are scales and techniques developed by the classroom teacher. By carefully following suggestions of experts in the area of measurement and with practice, the teacher

may become quite proficient in this skill. Some of the more usable and reliable techniques include Likert Scales, Semantic Differential Scales, Sociometric Techniques, Rating Scales, and Behavior Checklists" (1).

3. In the measurement of affective behavior, teacher observation and teacher judgment are also utilized. Teacher observation is described as a technique that systematically categorizes the behavior under consideration.
4. Teacher judgment can be utilized if the teacher constructs a rating scale or a checklist to be used in determining if the behaviors under consideration are being exhibited according to a given set of criteria.
5. Although teacher observation and judgment are common ways to measure affective behavior, subjectivity tends to be a very critical problem (1).*

(1) The Development and Evaluation of Behavioral Objectives, p. 69.

(1) As above, p. 67.

* See Discussion Question D in Part III.

C. Measuring Instruments for the Psychomotor Domain

1. According to Armstrong et al., "Measurements in the cognitive and affective domains of behavior assess what might be called internal behaviors. However, since the psychomotor domain deals primarily with external behaviors, the measuring techniques differ in some respects from those used in the cognitive and affective domains. Basically, this difference involves how the responses to the measuring instrument are recorded. Rather

than having the individual respond to the instrument directly as in the cognitive and affective domains, usually another person is required to observe this individual performing the given psychomotor skills under consideration and then record the observed performance" (1).

2. Measuring instruments that are available in the psychomotor domain include observation systems, rating scales, and checklists.

(1) The Development and Evaluation of Behavioral Objectives, p. 80.

D. Types of Written Test Items

1. Basically, written test items can be classified as either objective or essay.
2. Objective questions present the learner with a very structured situation that limits the type of response he makes. He must either select the correct answer from several alternatives, supply the correct answer, or determine the truth or falsity of a given statement. Types of objective test items include:

   a. multiple-choice items;
   b. matching items;
   c. true-false items;
   d. completion items.

   In general, objective test items are easier to administer, quicker to score, and provide more objective results than essay items.

3. Essay questions, on the other hand, measure the student's ability to select, organize, and integrate ideas. According to Butler, "Because the goal of criterion testing is to determine whether the student can meet the requirements of the objective and not whether he can write extensively and well, essay items have little or no place in a criterion test. This fact, coupled with the need to score the tests objectively, should lead you to reject essay items in most cases" (5).

(5) Instructional Systems Development for Vocational and Technical Training, p. 199. For a more detailed discussion of the advantages and disadvantages of objective items (and the various types of objective items) and essay items see: (21) Individualizing Vocational and Technical Instruction, Chap. 6; (9) Designing Objective, Essay, and Performance Tests; and (8) Home Economics Evaluation, Chaps. 6-10.

E. The Performance Test

1. The assessment of student performance in vocational education may take two forms:
   a. the performance test, and
   b. product evaluation.
2. The Performance Test. A performance test is a test that requires a student to accomplish a job-like task under controlled conditions, controlled conditions meaning those that will give the student the best possible chance to display the skill the test is to measure, and those that do not change from one student to another (5).
   a. Performance tests are used when the instructor is interested in determining if the student can perform the correct

(5) As above, p. 127.

process and if he can produce the correct product.

b. The test requires the instructor to observe the student as he is completing the process; therefore, the instructor can also determine if the product is completed correctly.

c. Since these tests must be administered on a one-to-one basis, they present certain difficulties for the instructor, who cannot be responsive to other students in the class at the time of performance testing.*

* See Discussion Question E in Part III.

3. Product Evaluation. Product evaluation is used when the instructor is primarily interested in whether the student can produce the correct product. It doesn't allow the instructor to assess the process that created the product--that is, the correct product may be completed by a correct or an incorrect process. It does, however, free the instructor to evaluate the performance of students after class or during periods when students do not require assistance.

4. Basic procedures for developing a performance evaluation instrument are:

   a. Specify the objective.

   b. Determine if you want to evaluate the performance with a performance test or a product evaluation.

c. If a performance test is used, list the procedural steps. If a product evaluation is used, list the points to be observed after the performance is completed. (Make sure that the steps or points are independent, that each contains only one performance, that each begins with a verb indicating the behavior expected of the student, and that all steps are listed.)

d. Identify critical items.

e. Determine if you need instructor checkpoints when using a product evaluation.

f. Determine the criteria for judging satisfactory completion of each step.

g. Establish the acceptable mastery level score for the instrument (21).

(21) Individualizing Vocational and Technical Instruction, p. 185.

F. Study Activities

Based on your reading of the content outline and any additional references as suggested, complete the following activities.

Techniques for Assessing Instructional Objectives in the Domains of Learning (22)

The following material enumerates some techniques that might be useful in assessing student achievement of instructional objectives in the three domains of learning: cognitive, affective, and psychomotor.

Assessment Techniques for the Cognitive Domain

The following techniques are appropriate for assessing student achievement of instructional objectives concerned with course content and factual information:

1. noting written or oral response to selected questions or issues listed in a pre-test or an exit test;
2. using teacher-made written tests consisting of objective-type questions;
3. having students prepare a short paper or essay with standards and criteria for assessment;
4. having a student chair or serve as a member of a committee, preparing and presenting a report on some aspect of a unit of instruction;
5. assessing a student's response to questions raised by an instructor in a group instruction review.

Assessment Techniques for the Affective Domain

The following technique is appropriate for assessing student achievement of instructional objectives concerned with the interests, attitudes, appreciations, and adjustments of the learner: using an attitude checklist that specifies behavioral criteria for judging student achievement of appropriate job-related attitudes.

Assessment Techniques for the Psychomotor Domain

The following techniques are appropriate for assessing student achievement of instructional objectives concerned with motor skills:

1. observing the student as he demonstrates a skill or the application of knowledge;
2. assessing a finished product that required the use of the psychomotor skills being assessed;
3. using a performance test in which the student demonstrates the psychomotor ability as part of the test.

1. Complete the following multiple-choice questions by marking an "X" by the specific learning domain being tested by the assessment technique described.

   a. As part of a performance test, an instructor observes a student to determine whether or not he is able to operate a power saw, following correct procedures.

      _____a. cognitive domain
      _____b. affective domain
      _____c. psychomotor domain

   b. A teacher prepares a series of multiple-choice questions to test the student's knowledge of the various types and uses of power saws.

      _____a. cognitive domain
      _____b. affective domain
      _____c. psychomotor domain

   c. With a list of characteristics that describe a safety conscious individual, an instructor observes a group of students working in a wood shop to determine whether or not these students are safety conscious.

      _____a. cognitive domain
      _____b. affective domain
      _____c. psychomotor domain

d. As part of a performance test, an instructor observes a group of retail sales trainees in a roleplaying session to determine whether or not they are able to establish good rapport with customers.

_____a. cognitive domain
_____b. affective domain
_____c. psychomotor domain

e. An instructor selects a standardized test consisting of matching items that determines whether or not dental assisting students are able to identify basic tools used by the dentist.

_____a. cognitive domain
_____b. affective domain
_____c. psychomotor domain

Types of Tests

In vocational education, there are two basic types of tests:

1. written tests, and
2. performance tests.

Written tests are designed to measure achievement of objectives primarily in the cognitive domain. Performance tests are designed to measure achievement of objectives primarily in the psychomotor domain. However, aspects of these tests may also be used to measure achievement of objectives in the affective domain. Their primary use is our major concern here.

Written tests consist of test questions that can be classified into two major groups:

1. objective test questions, and
2. subjective (essay) test questions.

Both basic types of tests, written tests and performance tests, may be either standardized or nonstandardized; that is, they may be purchased from a commercial test publisher, or they may be prepared by the individual instructor--teacher-made.

Keep these distinctions in mind as you proceed through this module.

Objective Tests and Subjective Tests: Advantages and Limitations (9)

1. An objective test is a type of test so designed that the score can be determined objectively and will be essentially the same regardless of who determines the score. Typical objective test questions include true-false, multiple choice, completion, matching, and pictorial recall. Objective tests may and should use more than one type of question. Each type of question has its own unique merits and limitations.

   a. True-False

      This type of test is generally inferior to other types since the element of "guessing" is always present. Remember, a person who knows absolutely nothing about the subject will average 50 percent correct by just answering all the questions. Furthermore, educators claim that even suggesting a negative answer is a poor practice in teaching.

      If true-false tests are used, there should be a relatively large number of questions, and there should be approximately an equal number of true questions and false questions. The student should be required to place a circle around the "T" or "F" in the corresponding right-hand column. For scoring, the number of incorrect answers may be subtracted from the number of correct answers.

      Advantages:

      (1) Comparatively easy to construct
      (2) May be applied to a wide range of subject matter
      (3) Objective and easy to score using a key
      (4) Permits a wide sampling of knowledge in a unit of work

Disadvantages:

(1) Includes negative suggestion

(2) Guessing factor is 50-50. Modifying corrects for this factor but the modification techniques usually confuse students

b. Multiple Choice

In this type of test the student must select the most appropriate answer from a minimum of four possible answers. Care should be taken to avoid more than one possible correct answer in the one-correct-answer type or more than one possible incorrect answer in the reverse multiple choice type.

Advantages:

(1) Tests judgment, reasoning, and discrimination of students

(2) Tests more than memory for factual knowledge (tests by recognition rather than recall)

(3) Very adaptable to who, what, when and where situations

(4) Reduces guessing factor from one-half to one-quarter

Disadvantages:

(1) practically none

(2) Initial construction of multiple choice items is time-consuming but this factor is offset by usefulness of questions

c. Completion

This type of test requires students to supply the answer to an incomplete statement or question by recalling one or two words, numbers, dates, or symbols. This type of testing requires that the student supply the exact answer intended. For example: "A letter that has a descender is _____." Avoid ambiguous statements.

Advantages:

(1) Tests memory
(2) Stimulates study habits
(3) Eliminates guessing

Disadvantages:

(1) Not a good measurement of student knowledge (emotional factors involved in test writing, i.e., fear, tension, nervousness)
(2) More difficult to score
(3) Measures only factual knowledge

d. .hing

A matching test is one that consists of matching words in one column with a closely related word or words in scrambled order in a second column. If for no other reason, they are used to add a certain amount of variety and interest to the otherwise boring task of taking a test.

Advantages:

(1) Comparatively easy to construct
(2) Objective and easy to score
(3) Efficient as a space and time saver
(4) When properly constructed, the guessing factor can be practically eliminated

Disadvantages:

(1) Inferior to multiple choice items for measuring judgment and application--apt to stress memorization of facts
(2) Unless properly constructed may include irrelevant clues to correct response
(3) Unless skillfully prepared may be time-consuming to student

e. Pictorial Recall

Identification tests, in which various parts of a drawing are to be identified, not only have an interest value, but are also quite effective for testing nomenclature for tasks, tools, materials, and parts of objects.

Advantages:

(1) Tests memory
(2) Stimulates study habits
(3) Eliminates guessing
(4) Easy to score

Disadvantages:

(1) Measures only factual knowledge
(2) Emotional factors are involved in writing tests, i.e., fear, tension, nervousness

2. A subjective test, such as an essay test, is one that is scored on the basis of the scorer's personal judgment of the worth of each answer. Essay type questions are fairly easy to prepare and are adaptable to most subjects and most classroom conditions. The chief disadvantage is that they are hard to score fairly. This is because grading is based chiefly on opinion, which may be influenced by neatness, "literary" ability rather than subject matter, or personality conflicts between the student and instructor.

Advantages:

(1) Measures student's ability to organize his/her thoughts and express himself/herself clearly
(2) Takes a comparatively short time to prepare

Disadvantages:

(1) Time consuming to score
(2) Difficult to score objectively
(3) Time-consuming for student to write
(4) Offers poor coverage of area to be tested
(5) Penalizes the student who is unable to express himself/herself well
(6) Lacks reliability

The Requirements of Good Tests (9)

A test is only as good as its results. In other words, if a test is "good," it is good because it accomplishes its purpose effectively and economically in a particular situation. Therefore, a good test is one that is objective, valid, reliable, comprehensive, and provides for economy of time in giving and scoring. Analyze these qualities carefully.

1. Objective

   When a test can be used by two or more examiners of equal competence and give identical or similar scores, it is said to have objectivity. It is a quality dependent on purely impersonal, factual evidence rather than on judgment, personal opinion, or bias. Objectivity, therefore, applies to the giving and scoring of a test and not to the person taking the test.

2. Valid

   When a test measures what it is intended to measure, it is said to have validity. Validity requires careful selection of test items to avoid irrelevant and nonessential questions that are not true measures of knowledge or ability. Every term in the test should be representative of the main purpose of the unit of study being tested.

3. Reliable

A test is said to have reliability when it gives consistent results whether given at different intervals to the same group or given to different groups who have received the same instruction. Reliability, therefore, refers to the accuracy with which a test measures the things that it is supposed to measure.

4. Comprehensive

A test should provide adequate coverage of the subject or that part of the subject to be tested. The questions should cover all the points emphasized in the lesson. Written tests, such as the old essay type, have only a few questions and are obviously not comprehensive.

5. Convenient

A test should be easy to use and should provide for economy of time in administering and scoring. Its construction should be such that it is possible to test a larger number of items in a class period, and the instructor is able to score a larger number of tests with true objectivity.

2. Having read the preceding material on "Types of Tests," "Objective Tests and Subjective Tests: Advantages and Limitations," and "The Requirements of Good Tests," fill in the chart on the following page, indicating the advantages and limitations of objective tests and subjective tests for the criteria indicated.

CRITERION	Objective Tests		Subjective Tests	
	Advantages	Limitations	Advantages	Limitations
1. Objectivity				
2. Validity				
3. Reliability				

CRITERION	Objective Tests		Subjective Tests	
	Advantages	Limitations	Advantages	Limitations
4. Comprehen-siveness				
5. Convenience				

3. In Module 7, Derivation and Specification of Instructional Objectives, you wrote instructional objectives for specfic occupational tasks. (See the Objectives Specification Sheet you completed on page 53 of the Study Guide for Module 7.) In Module 8, Development of Instructional Materials, you selected instructional strategies for accomplishing these objectives. (See the Selection of Instructional Strategies forms you completed on page 30 of the Study Guide for Module 8.) Now you will have the opportunity to select assessment approaches/techniques for the objectives you specified in Module 7. Be sure you have objectives representing the three domains of learning: cognitive, affective, and psychomotor.

   Look at each of your objectives and select an approach/technique to assess students' mastery of that objective. Use a form like the one provided on the next page to write the objective and the corresponding assessment technique. (You will need to prepare a form for each objective. Adequate space is not provided in this guide, so use additional sheets of paper as necessary.)

SELECTION OF ASSESSMENT APPROACHES/TECHNIQUES

Objective:

Domain:

Assessment Approach/Technique:

Content Outline	Activities-Resources

Goal 9.3: Develop An Evaluation Plan and Construct Test Instruments for Measuring Student Achievement of Instructional Objectives.

A. Test Construction

Test construction is a highly specialized subject, and many texts exist on the subject. The important point here is that good test construction is absolutely essential to the success of instructional materials development. The entire developmental process can stand or fall on the quality of the criterion tests.

B. Implementing Criterion-Referenced Measurement

To summarize this module on criterion-referenced testing, there are three essential steps to remember in implementing criterion-referenced measurement. They are:

1. Prior to instruction, prepare a set of instructional objectives for the unit of instruction.
2. Select appropriate assessment approaches/ techniques to assess students' mastery of the stated objectives. The appropriateness of the technique selected will largely be determined by the nature of the skill or competency specified in the objectives.
3. Match particular assessment techniques

selected with the performance behaviors delineated in the instructional objectives. It is important that the specific test item correspond to the level and type of behavior specified in the objective. Perhaps the single most effective method of ensuring precision and accuracy in criterion-referenced measurement is the careful matching of test items to performance objectives (14).

(14) Objectives for Instruction and Evaluation, p. 118.

C. Wrapup of Module* *

*See Classroom Activity 4 in Part III.

*See Discussion Question F in Part III.

D. Study Activities

Based on your reading of the content outline and any additional references as suggested, complete the following activities.

1. Read pp. 40-43 ("Criterion Examination") of Mager and Beach, Developing Vocational Instruction. Then complete the following activity.

   For each of the objectives provided in this exercise, there is a list of possible test items. Indicate by writing "yes" or "no" whether or not that test item is appropriate for assessing the objective.

   a. OBJECTIVE: When approached by a prospective customer, respond in a a positive manner (with a smile, a suitable greeting, and pleasant tone of voice).

      _____a. Describe the three basic characteristics of a positive response to the approach of a prospective customer.

      _____b. Look at the following ten photographs and write the number of those that represent a correct response to the approach of a prospective customer.

      _____c. Watch the following ten film clips and write down the number of those that represent a correct response to the approach of a prospective customer.

      _____d. When the instructor hangs the "customer" sign around his neck and approaches you, make the correct response to the approach of a prospective customer.

      _____e. Write a paragraph describing the importance of each element of the response to customer approach.

      _____f. When approached by each of five students selected by the instructor, make the appropriate response to customer approach.

b. OBJECTIVE: Be able to type a business letter in accordance with standards described in Company Manual 12-21.

_____a. Describe the five basic elements of a business letter.

_____b. Sort the ten sample letters into piles representing those that are written in accordance with Company standards and those that are not.

_____c. On the five sample letters given, circle any errors or items not in accordance with Company standards.

_____d. Describe in a paragraph the rationale for the business letter standards currently in effect.

_____e. From the rough copy given, type a business letter in the form set out by Manual 12-21.

_____f. Tell how you would instruct a secretary in the preparation of business letters according to current policy.

c. OBJECTIVE: Be able to read a domestic electric power meter correctly to the nearest unit and record it on the appropriate page of the Meter-Reader's log.

_____a Define kilowatt-hour.

_____b. Of the five dials on the domestic meter, which records "thousands of units"?

_____c. Look at this picture of a dial. What is the reading?

_____d. Look at the dials on these domestic meters. What are the readings?

_____e. Record on the appropriate page of your log the readings of each of these ten domestic meters.

d. OBJECTIVE: Be able to construct a parallelogram.

_____a. Define parallelogram.

_____b. Describe the difference between a parallelogram and a rectangle.

_____c. Look at the following figures and draw a circle around the parallelograms.

_____d. Draw a parallelogram whose sides are 1" and 3" in length.

2. Read the Weber and Lucas article, "Evaluating Student Progress," in The Individual and His Education. Then complete the activity below.

As a final activity in Module 8, Development of Instructional Materials, you developed a lesson plan for a unit of instruction. (See Goal 8.2, p. 69, in the Study Guide for that module.) Now you will have the opportunity to develop an evaluation plan for assessing achievement of the objectives for that unit of instruction.

Using the Weber and Lucas article as a guide, develop a Table of Specifications for your unit of instruction. This table will serve as a plan for test development. Indicate the content of your unit of instruction and the various dimensions of your instructional objectives.

When you have completed the Table of Specifications, look at it to see if it represents a balanced evaluation scheme. Then answer the following questions.

a. Does your Table of Specifications represent a balanced evaluation scheme? If not, how would you explain this? Perhaps there is a good reason.

b. If your Table of Specifications is not balanced and there doesn't seem to be good reason for this, how would you change the content and objectives for your unit of instruction in order to create a more balanced unit?

3. Having read the Weber and Lucas article, "Evaluating Student Progress," and having constructed an evaluation plan for your unit of instruction, you will now have an opportunity to actually construct the test instruments for the unit.

Develop written tests, performance tests, and tests to measure attitudes, as appropriate. The important concern is to develop test items that are appropriate for assessing achievement of your objectives.

Many texts on principles of test item construction exist and several are mentioned in the Weber and Lucas article. Feel free to use whatever of these reference materials you find necessary to develop test items. (Adequate space is not provided in this guide, so use additional sheets of paper as necessary.)

Part III:

Group and Classroom Activities

PART III
GROUP AND CLASSROOM ACTIVITIES

Classroom Activities

NOTE: The following activities are designed to stimulate discussion in the classroom on specific topics covered in this module. The activities are designed to be used after student self-study; however, depending on the background and abilities of students, these activities may not require previous self-study. All classroom activities are keyed to the Content Outline to indicate an appropriate point for participation.

1. Debate the following issue:

   Norbert Wiener noted that the human brain is able to handle value ideas--ideas that are not quantifiable and that any computer would have to reject as formless. (Norbert Wiener, God and Golem, Inc. Cambridge, Mass.: The M.I.T. Press, 1964, p. 73.) Yet some proponents of behavioral objectives contend that the teacher is not engaged in instruction when dealing with objectives that are not describable in terms that car be quantifiably measured. (W. James Popham and Eva L. Baker, Systematic Instruction. Englewood Cliffs, New Jersey: Prentice-Hall, Inc., 1970, p. 141.) (26)

   Students should divide into two teams, one team representing the Wiener point of view, and one team representing the Popham and Baker point of view on this issue of measuring learning outcomes in connection with instructional objectives. If there are students in the class that do not feel strongly about either point of view, they should form a third team representing a point of view midway between the extremes.

2. Discuss the following issue: A contemporary educational critic contends that existing practice "makes it clear to students that the purpose of testing is not evaluation but rating--to produce grades that enable the school to rank students and sort them in various ways for administrative purposes. The result is to destroy any interest in learning. . ." (Charles E. Silberman, Crisis in the Classroom. New York: Random House, 1970, p. 348.) Do you agree with this statement? Explain the reasons for your point of view. (26)

   Students should provide examples from any vocational courses they have taken that support their point of view on this issue.

3. Students should group themselves into two teams to debate the issue of whether or not instructors or curriculum specialists can more effectively acquire a set of measurable and appropriate objectives with corresponding test items by generating their own or by selecting them from other sources.

   Consider such practical matters as:
   Are other sources of measurable objectives with corresponding test items available? If so, are these objectives and test items suitable for the local situation?
   Are instructors or curriculum specialists likely to have time to generate their own objectives and test items? If not, is it possible that a local curriculum team might effectively function to develop objectives and test items?

4. As a wrapup activity for this module, select any learning module (teacher-made or commercial) that may be available in the classroom. Students should analyze the module for match of objectives and test items. Based on how well objectives and test items match, students should determine whether or not they would recommend the module for further use.

Activities for Additional Credit

NOTE: These activities are designed for the student who wishes to obtain additional credit beyond the basic requirements of this module. You may choose to write a paper on one of these activities, or discuss the activity with the instructor, or you may select some other method to complete the activity.

1. Examine some teacher-made tests in your area of specialization and evaluate each item according to the cognitive levels represented in Bloom's Taxonomy of Educational Objectives. What cognitive levels receive the greatest emphasis? What cognitive levels are totally neglected? Can you devise some test items that represent the cognitive levels of application, analysis, synthesis, and evaluation?

2. Do the same for the affective domain as you did in Item 1 above.

3. In Silvius and Bohn, Planning and Organizing Instruction (22), the authors recommend the following book for procedures to develop criteria for measuring student achievement of instructional objectives in the affective domain: Robert F. Mager, Goal Analysis (Belmont, California: Fearon Publishers, 1972). Read this book--it's another Mager shortie--and using the procedures presented there, describe the specific performances that would indicate achievement of several of the affective objectives from the unit of instruction you developed for this group of modules. When you have listed these specific performances, develop criterion test items that indicate whether or not the student has achieved the affective objective.

4. Visit two vocational classes in your area of specialization and describe the types of measurement being used to assess student achievement. For what purposes are these types of measurement being used? Is the use appropriate for the purpose? What recommendations would you make to improve the methods of assessing student achievement?

5. In Volume I of his Educational Psychology, E. L. Thorndike (1913) described the problems of judging student achievement in relation to other students' achievement. Cite arguments from Thorndike that would support today's concept of "criterion-referenced" measurement as a means to assess student achievement.

6. It is possible that someday, upon designation of instructional objectives, a computer could devise a table of specifications (evaluation plan), select test items from a test-item pool, print the test, score and grade it, and analyze the results. What do you see as the possible advantages and disadvantages of such a criterion-referenced measurement system?

7. By researching the literature, or by any other means of your choosing, locate an example of a testing program in any area of vocational education that uses both norm-referenced and criterion-referenced tests. Then in a 3 to 5 page paper, summarize this program, highlighting the appropriate and effective uses of norm-referenced measurement and criterion-referenced measurement. Use the Ocean View program described in the Special Report: The Nature and Uses of Criterion-Referenced and Norm-Referenced Achievement Tests (page 31 of the Study Guide) as a guide for the preparation of your paper.

Part V:

Appendices

PART V
APPENDICES

Appendix A:
Possible Study Activity Responses

GOAL 9.1

1a. b
b. a
c. b
d. b

2a. a
b. b
c. b
d. a

3a. Psychological tests were administered to thousands of Army recruits during World War I. These tests, which employed the concept of the IQ or "mental ratio," caught on, the idea spread, and the form was set. When the war was over, the schema of the mental test, invented to discover and predict aptitude, was remodeled for school use--not only for this purpose but also to test school achievement for diagnostic and training purposes. Standardized subject-matter tests and test batteries multiplied. (20)

b. Robert Glaser coined the term "criterion-referenced measurement" in a 1963 article in the American Psychologist, called "Instructional Technology and the Measurement of Learning Outcomes: Some Questions."

c. Probably the programmed instruction movement of the early 1960s with its emphasis on behavioral objectives gave the greatest impetus to criterion-referenced measurement (which measures individual achievement of objectives).

d. In large part, achievement measures currently employed in education are norm-referenced. This emphasis upon norm-referenced measures has been brought about by the preoccupation of test theory with aptitude, and with selection and prediction problems; norm-referenced measures are useful for this kind of work. However, the imposition of this kind of thinking on the purposes of achievement measurement raises some questions (20).

e. Glaser's primary concern regarding the measurement of learning outcomes is how to assess existing levels of competence and achievement and the conditions that produce them.

4a. It is in the technical sense that criterion-referenced tests are rather new. Only in the last few years have those concerned with the theoretical and technical issues of educational and psychological measurement undertaken any sustained large-scale effort to deal with criterion-referenced tests. It has been even more recently that publishers of tests for schools have tried to offer any substantial tests of this sort. Therefore, consensus among the experts about technical requirements for the tests has yet to be reached and little about these matters can be found in courses and textbooks.

b. Other current terms in use for criterion-referenced tests include: "domain-referenced," "objectives-based."

c. There is little unanimity of opinion about the meaning of the word "standardized" as a descriptor of tests. If "standardized" is defined as describing a test that has explicit instructions for administration and that was constructed to meet technical standards, then a criterion-referenced test can be considered standardized.

5a. b

b. a

c. a

d. b

e. a

f. b

6a. VARIABILITY

CRM: Variability is irrelevant; it is not a necessary conditon for a good criterion-referenced test.

NRM: Variability is essential; since the meaningfulness of a norm-referenced score is basically dependent on the relative position of the score in comparison with other scores, the more variability in the scores the better.

b. ITEM CONSTRUCTION

CRM: Criterion-referenced item writers are guided by the goal of making sure the item is an accurate reflection of the criterion behavior.

NRM: Norm-referenced item writers are guided by the goal of developing items that produce variability.

c. RELIABILITY

CRM: Criterion-referenced tests should be reliable, that is, they should be internally consistent. However, it is not obvious how to assess the internal consistency; the classical procedures are not appropriate because they are dependent on score variability. Every student could obtain a perfect score on a criterion-referenced test, yet by classical standards this test would not be considered internally consistent.

NRM: Norm-referenced tests should be reliable, that is, they should be internally consistent. Classical procedures for assessing internal consistency are appropriate for norm-referenced measurement.

d. VALIDITY

CRM: Criterion-referenced measures are validated primarily in terms of the adequacy with which they represent the criterion.

NRM: Many of the procedures for assessing the validity of norm-referenced tests are based on correlations and thus on variability.

e. ITEM ANALYSIS

CRM: For criterion-referenced tests, the use of discrimination indices (item analysis procedures that identify those items that do not properly discriminate among individuals taking the test) must be modified. An item that does not discriminate need not be eliminated.

NRM: Item analysis procedures have traditionally been used with norm-referenced tests to identify those items that do not properly discriminate among individuals taking the test. If an item does not properly discriminate between the more and less knowledgeable learners, the item should be eliminated.

f. REPORTING AND INTERPRETATION

CRM: When interpreting an individual's performance on a criterion-referenced test, group-relative indices are not appropriate. The individual has either mastered the criterion or he has not. In reporting an individual's performance, one alternative is the use of an "on-off" approach; the student either has or has not achieved the criterion.

NRM: In interpreting the results of an individual's performance on a norm-referenced test, the concern is with the individual's performance in relation to the performance of other individuals.

GOAL 9.2

1a. c
b. a
c. b
d. b
e. a

2. (See following pages)

CRITERION	Objective Tests		Subjective Tests	
	Advantages	Limitations	Advantages	Limitations
1. Objectivity	This the major advantage of objective tests.	---------	---------	This is the major limitation of subjective tests.
2. Validity	If objective test items are tied directly to instructional objectives, validity can be a major advantage of objective tests.	If objective test items are not tied directly to instructional objectives, then one can question the validity of these items.	If standards and criteria for assessment are specified, then a subjective test can be valid.	If standards and criteria for assessment are not specified, then one can question the validity of these items.
3. Reliability	If test items are well-written and precise, then reliability can be a major advantage.	If test items are vaguely worded and open to interpretation, then they are not reliable.	---------	Subjective tests often lack reliability.

Discussion Questions

A. What is the difference between measurement and evaluation?

(For many years educators in this country have tossed around the term "evaluation" with almost indifferent imprecision. For some, the expression referred exclusively to the grading operations wherein pupils were assigned A, B, C, etc. To others, it meant essentially the same as "measurement." Still others thought of evaluation as experiments to discover if Method A was better than Method B. Although each of these notions of educational evaluation has been subscribed to by many, each is clearly inconsistent with the conception of educational evaluation endorsed by most educational leaders today.) (19)

B. What educational phenomena of the 1960s might have given impetus to criterion-referenced measurement?

(Probably the programmed instruction movement of the early 1960s with its emphasis on measurable instructional objectives gave the greatest impetus to criterion-referenced measurement--which measures student achievement of instructional objectives.)

C. What advantages, limitations, and/or dangers do you see in the position that teachers must specify and measure all instructional objectives?

(This is a matter of great debate in educational circles. Proponents of instructional objectives hold that the only sensible reason for the educator's engaging in instruction is to modify the learner's behavior; therefore, these intended changes must be described in terms of measurable learner behaviors. On the other hand, educators who do not feel the need to specify all learner behavior in terms of instructional objectives believe that the problem with excessive insistence on building specifications for each and every instructional objective is that human beings are not built like automobiles or washing machines. The consequence of such detailed specifications in education, they feel, is that achievement comes to denote the sort of thing that a well-planned machine can do better than a human being can, and the main effect of education, the achieving of a life of rich significance, drops by the wayside.) (26)

D. Many educators acknowledge that affective objectives and their measurement have not received adequate emphasis in the curriculum, particularly the vocational curriculum. How do you account for this?

(Some of the major reasons for neglect of affective objectives in education include:

1. The failure of teachers and curriculum specialists to appreciate fully the necessary interrelationships and interdependence between affective and cognitive objectives;

2. the concentration on cognitive learnings in the curriculum reform projects of the 1950s and 1960s;
3. the emphasis on cognitive learnings--particularly those at the lower levels of the taxonomy--in traditional schooling, coupled with the long-held concept of mind as an entity separate from the emotions, along with the persistent, time-worn belief that the mind is best strengthened through rigorous intellectual exercise;
4. the new emphasis on operant conditioning and the treatment of the learner as an automatic mechanism;
5. the enormous difficulty inherent in teaching for and evaluating affective learning in conjunction with cognitive learning;
6. the controversies attached to evaluating attitudes, feelings, emotions, and values.) (26)

E. Discuss the relation of performance tests to written tests. When might a written test be considered a performance test? Many instructors speak of written tests and performance tests as if they were separate and distinct types. Written tests are often considered poor measures of proficiency, while performance tests are thought to constitute the only real measures of performance. Is this necessarily true?

(A performance test is a test which requires a student to accomplish a job-like task under controlled conditions. Because some written tests are indeed job-like, they too can be considered performance tests. On the other hand, just because a test item involves equipment and requires the student to perform something does not mean that it is job-related. Also, the fact that performance is involved does not assure accurate measurement of ability. The real requirement is that the test situation make demands of the student that are as similar as possible to those of the job. It is <u>what</u> is <u>measured</u> that counts in a performance test--not the procedure by which it is measured.) (5)

F. Do you have any problems or concerns regarding this module?

Part IV:

Student Self-Check

PART IV
STUDENT SELF-CHECK

Part A: Knowledge Assessment

GOAL 9.1

1. What are the definitions of the following items? (9.11)
   a. educational evaluation:
   b. educational measurement:
   c. criterion-referenced testing:
   d. norm-referenced testing:

2. What factors in American education contributed to an increasing emphasis on criterion-referenced measurement? (9.12)

3. Which type of test has as its major purpose the determination of the student's relative position within a group of students? (9.13)
   _____a. criterion-referenced test
   _____b. norm-referenced test

4. Which type of test consists of test items that are constructed to measure a predetermined level of proficiency? (9.13)
   _____a. criterion-referenced test
   _____b. norm-refererced test

5. State the difference between norm-referenced measurement and criterion-referenced measurement on the basis of: variability, reliability, and validity. (9.14)
   a. Variability:
   b. Reliability:
   c. Validity:

GOAL 9.2

6. As part of a performance test, an instructor observes a student to determine whether or not he is able to clean and make a hospital bed, following correct hospital procedure. What learning domain is this technique primarily assessing? (9.21)

   _____a. psychomotor domain

   _____b. cognitive domain

   _____c. affective domain

7. What are the two basic types of test questions used in paper-and-pencil tests? (9.22)

8. State the differences between objective tests and subjective tests on the bases of: reliability and comprehensiveness. (9.22)

   a. Reliability:

   b. Comprehensiveness:

9. What is a "performance test"? (9.23)

GOAL 9.3

10. Which of the following test items is appropriate for assessing this objective: Be able to recognize when a torch flame is appropriate for cutting half-inch steel? (9.31)

    _____a. Describe the characteristics of a torch flame that is appropriate for cutting half-inch steel.

    _____b. Look at the following eight color slides of good and bad flames and write the number of those appropriate for cutting half-inch steel.

    _____c. Given a welding torch, adjust the flame until it is appropriate for cutting half-inch steel.

    _____d. Tell how you would adjust the flame of a welding torch to make it appropriate for cutting half-inch steel.

Part B: Performance Assessment

The purpose of this part of the test is to assess your ability to perform some of the actual steps involved in the construction of criterion-referenced test instruments. You should complete it outside of class and use any reference materials that may be helpful. Be sure to have the materials you developed for the Performance Assessment portions of the Module 7 and Module 8 Self-Checks. You will now have a chance to build on the materials you completed there.

This test consists of completing each of the following items in order. As you finish each item, check it off and continue to the next. If you find any of the forms suggested in the Study Guide helpful in completing these steps, use them. Otherwise, you may use your own particular forms, as long as you complete each step below as indicated.

_____1. Select approaches/techniques for assessing the objectives of the two units of instruction for which you developed lesson plans in the Module 8 Performance Assessment. This should be a general listing of several possible approaches to assessing the objectives. Later, you will select one of these approaches and construct test instruments following the approach you selected. (9.24)

_____2. In the Module 8 Performance Assessment, you developed lesson plans for two units of instruction: one unit primarily in the cognitive or affective domain, and the other unit in the psychomotor domain. Now you are to develop an evaluation plan for assessing achievement of the objectives for those units of instruction. Using the Weber and Lucas article in The Individual and His Education as an aid, develop a Table of Specifications for each unit of instruction. These tables will serve as a plan for test development. Indicate the content of your instructional units and the various dimensions of your instructional objectives. (9.32)

_____3. From the list of possible approaches/techniques for assessing objectives that you developed for Item 1, select a final approach and construct criterion-referenced test instruments for your two units of instruction. Develop written tests, performance tests, and tests to measure attitudes, as appropriate. The important concern is to develop test items that match the specifications of your objectives. (9.33)

CRITERION	Objective Tests		Subjective Tests	
	Advantages	Limitations	Advantages	Limitations
4. Comprehensiveness	Ojective test items permit a wide sampling of knowledge, therefore comprehensiveness is a major advantage.	--------	----------	This is another major limitation of subjective tests.
5. Convenience	Objective tests are easy to administer and score.	Writing objective tests is time-consuming.	Writing subjective tests takes a relatively short time.	Subjective tests are time-consuming to score.

3. (The specific response to this activity depends on the particular objectives you specified in Module 7. In general, the approaches/techniques you select for measuring achievement of these objectives should be appropriate for assessing the specific behavior described by the objectives. For example, if an objective specifies that a student should be able to demonstrate some job-related task, the assessment approach should require the student to actually demonstrate that task, not to take a written multiple-choice test.)

GOAL 9.3

1a. a. no
b. no
c. no
d. yes
e. no
f. yes

b. a. no
b. no
c. no
d. no
e. yes
f. no

c. a. no
b. no
c. no
d. no
e. yes

d. a. no
b. no
c. no
d. yes

2. (The specific response to this activity depends on the particular objectives you specified for your particular unit of instruction. If your Table of Specifications does not represent a balanced evaluation scheme, there may be good reason. Perhaps the subject of your course emphasizes objectives of one type. For example, a course in retail arithmetic emphasizes application.)

3. (The specific response to this activity depends on the particular objectives you developed for your particular unit of instruction. The important concern is that your test items assess the behavior specified by the objectives. Check with your instructor.)

Appendix B:
Possible Self-Check Responses

Part A: Knowledge Assessment

GOAL 9.1

1. What are the definitions of the following terms? (9.11)

   a. educational evaluation:

      The determination of the worth of educational phenomena; the term generally refers to the evaluation of an educational enterprise, such as an instructional sequence, not to the evaluation of students within that enterprise.

   b. educational measurement:

      The assessment of the current status of an educational phenomenon in a precise fashion--that is, counting or enumerating so that the phenomenon can be more accurately described--without placing value (goodness or badness) on the phenomenon thus described.

   c. criterion-referenced testing:

      A form of educational measurement that ascertains an individual's status with respect to some criterion or performance standard.

   d. norm-referenced testing:

      A form of educational measurement that ascertains an individual's performance in relationship to the performance of other individuals on the same measuring device.

2. What factors in American education contributed to an increasing emphasis on criterion-referenced measurement? (9.12)

Probably the programmed instruction movement of the early 1960s with its emphasis on behavioral objectives gave the greatest impetus to criterion-referenced measurement.

The appearance of Robert Glaser's 1963 article in the American Psychologist, "Instructional Technology and the Measurement of Learning Outcomes: Some Questions," in which the term "criterion-referenced" measurement appeared for the first time, also drew attention to CRM.

3. Which type of test has as its major purpose the determination of the student's relative position within a group of students? (9.13)

_____a. criterion-referenced test

__X__b. norm-referenced test

4. Which type of test consists of test items that are constructed to measure a predetermined level of proficiency? (9.13)

__X__a. criterion-referenced test

_____b. norm-referenced test

5. State the difference between norm-referenced measurement and criterion-referenced measurement on the basis of: variability, reliability, and validity. (9.14)

a. Variability:

With norm-referenced measurement, the more variability in the test scores the better, since the meaningfulness of the score is basically dependent on the relative position of the score in comparison with other scores.

Variability is not a necessary condition for a good criterion-referenced test; in fact, variability is irrelevant since the meaning of the score is not dependent on comparison with other scores.

b. Reliability:

A norm-referenced test is reliable when all the items in it "measure the same thing" to some minimal extent, that is, when the test is internally consistent. Classical procedures to assess internal consistency are dependent on score variability and thus are appropriate only for norm-referenced tests.

Obviously, criterion-referenced tests should be internally consistent, but it is not obvious how to assess this. The classical procedures are not appropriate, and indices to assess internal consistency of criterion-referenced tests have not yet been developed.

c. Validity:

For norm-referenced tests, the assessment of their validity is based on correlations and thus on variability.

For criterion-referenced tests, the assessment of their validity is based primarily on the adequacy with which they represent the criteria.

GOAL 9.2

6. As part of a performance test, an instructor observes a student to determine whether or not he is able to clean and make a hospital bed, following correct hospital procedure. What learning domain is this technique primarily assessing? (9.21)

X a. psychomotor domain

b. cognitive domain

c. affective domain

7. What are the two basic types of test questions used in paper-and-pencil tests? (9.22)

   1. objective test questions
   2. subjective test questions

8. State the differences between objective tests and subjective tests on the bases of: reliability and comprehensiveness. (9.22)

   a. Reliability:

      Reliability is the greatest advantage of objective tests since test scoring is void of teacher bias.

      In most cases, the reliability of subjective tests is very low because scoring depends on the individual teacher evaluating the test, thereby opening the door for individual bias or prejudice.

   b. Comprehensiveness:

      Since objective tests permit a wide sampling of knowledge, comprehensiveness is a major advantage. With a test item for every objective, a criterion-referenced objective test will necessarily give comprehensive coverage of all desired behaviors.

      Comprehensiveness is a major limitation of subjective tests.

9. What is a "performance test"? (9.23)

   A performance test is a test that evaluates, under realistic conditions, the performance of tasks that have value in some life situation.

GOAL 9.3

10. Which of the following test items is appropriate for assessing this objective: Be able to recognize when a torch flame is appropriate for cutting half-inch steel? (9.31)

    ____a. Describe the characteristics of a torch flame that is appropriate for cutting half-inch steel.

    _X_ b. Look at the following eight color slides of good and bad flames and write the number of those appropriate for cutting half-inch steel.

    ____c. Given a welding torch, adjust the flame until it is appropriate for cutting half-inch steel.

    ____d. Tell how you would adjust the flame of a welding torch to make it appropriate for cutting half-inch steel.

(NOTE: Although the actual flame is probably more relevant for testing than color slides, it is less practical for discrimination training since it would take the instructor considerable time to misadjust a flame to present the student with a predesigned array of stimuli.)

Part B: Performance Assessment

In scoring Part B, you should be primarily concerned with the techniques and processes used to construct test instruments and with the appropriateness of test items for assessing achievement of specific objectives. Your personal judgment will be a major factor in scoring Part B. However, for the test items indicated below, assessment should consider specific factors:

Item 1. The approaches/techniques selected for measuring achievement of the objectives should be appropriate for assessing the specific behavior described by the objectives. For example, if an objective specifies that a student should be able to demonstrate some job-related task, the assessment approach used should require the student to actually demonstrate that task, not to take a written multiple-choice test.

Item 2. Because a Table of Specifications will be developed for instructional units primarily representing one learning domain (cognitive, affective, or psychomotor), the evaluation plan will emphasize test items

in that particular domain. This emphasis is reasonable, and one should not be downgraded for having an "unbalanced" plan.

Item 3. The test items developed should match the specifications of the objectives, that is, the specific behavior described by the objective should be the specific behavior that the test item assesses.

☆GPO 692-192-1976