ESSENTIAL SKILLS FOR COLLEGE TEACHING

AN INSTRUCTIONAL SYSTEMS APPROACH

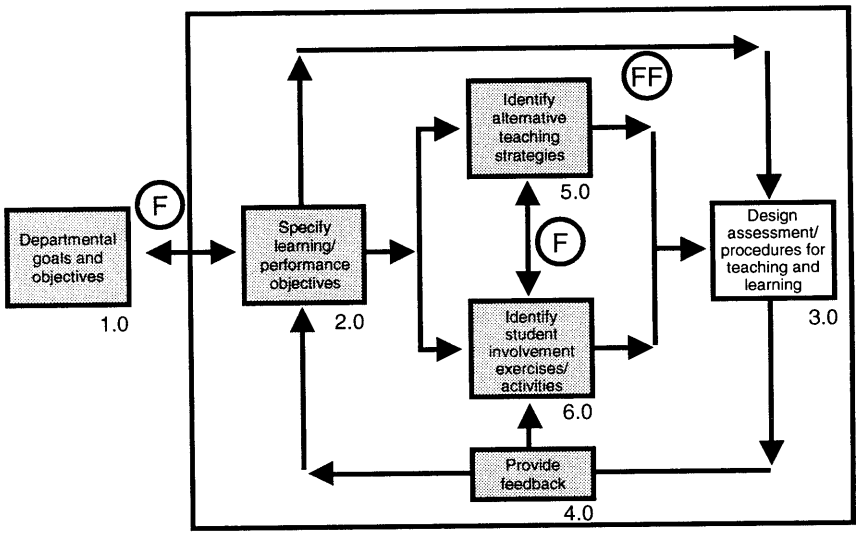
third edition

Thomas E. Cyrs

New Mexico State University

Chapter 19

Constructing Valid Tests to Match Your Learning Objectives



Special thanks to Alvin Kent, former Director of Media Services at Iowa State University, who originally co-authored the material in this chapter for another project.

Joint Statement on Rights and Freedoms of Students

on

Protection Against Improper Academic Evaluation

by

American Association of University Professors
U.S. National Student Association
Association of American Colleges
National Association of Student Personnel Administrators
National Association of Women Deans and Counselors

Students should have protection through orderly procedures against prejudiced or capricious academic evaluation. At the same time, they are responsible for maintaining standards of academic performance established for each course in which they are enrolled.

(Milton and Edgerly, 1976)

CONSTRUCTING VALID TESTS

LEARNING PERFORMANCE OBJECTIVES

Using the learning performance objectives from a section of your course, identify the most valid assessment procedures. Justify to the class why they were valid.

KEY IDEAS

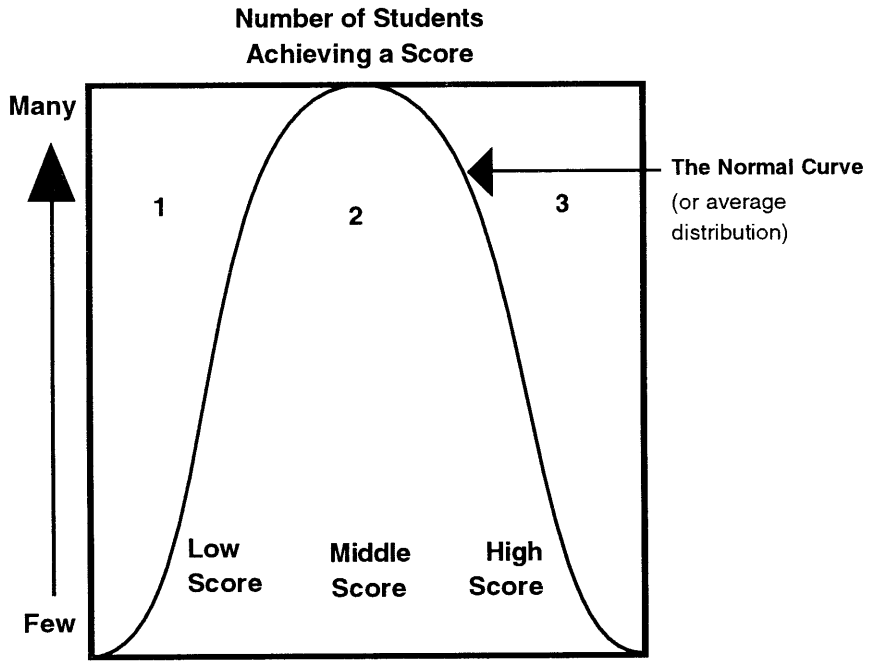
1. All testing techniques, performance on paper and pencil, must be based on the same cognitive level as the learning performance objective and the same class of behavior to be demonstrated.
2. Criterion-referenced tests are absolute measures of performance rather than relative measures as demonstrated in norm-referenced.
3. Instruction can be qualitatively improved through testing.

NEW VOCABULARY

- valid tests and reliable tests
- criterion-referenced
- content validity
- predictive validity
- essay tests
- norm-referenced
- normal curve
- construct validity
- performance test

One of the most significant issues in education is test validity–did the test measure what it was intended to measure. A test, regardless of the type, is an instrument to systematically measure a sample of student performance. It is systematic in the way that it generates a sample of performance from a domain of performance, in methods of scoring, and in interpreting results. Performance includes skills, knowledge, and attitudes.

The selection of a test procedure must be based on the type of performance required of the student. The learning performance objectives that define specifically



CRITERION-REFERENCED TESTING

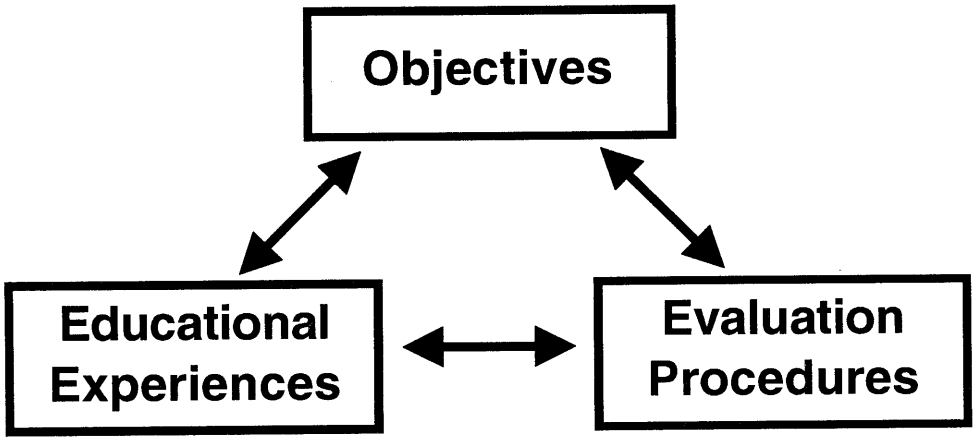
Criterion-referenced testing is relatively new and is based on the assumption that mastery of certain performances is critical to the future success of the student. Mastery of certain skills may also have critical consequences for "safe" performance in such areas as the health professions and engineering. Criterion-referenced or mastery learning is based on the research of Carroll (1963) and Bloom (1968). Once a mastery level is determined, comparative data is useless. The most critical question posed in this approach is, *"To what degree did the student meet the required level of performance as specified in the learning performance objectives?"* This approach considers the standard of performance as the criterion against which to judge someone. Can a surgeon get a less than competent grade, a "B-," in open heart surgery or say that he or she did not learn that piece of anatomy? How would you like to fly in a plane in which the pilot got a "B" in takeoff and a "C" in landing? Would you like to cross a bridge when told that the bridge engineer received an average of "B-" in bridge-building?

Criterion-referenced tests are designed to *minimize* variance. Evaluation is based on how closely the student comes to achieving the performance objective.

The student's achievement is compared with his/her *own* progress. Test items are designed to discriminate between the student's ability or lack of ability to reach a designated level of performance. High achievement translates into virtually all students reaching a minimum level of performance.

In summary, the option for testing that is chosen will interact and reflect upon the approach to teaching. The major questions addressed in these two approaches to testing are very important. *Norm-referenced testing* asks, "What proportion

what the student is expected to do form the framework for the development of testing situations. The experiences that students have in the learning activities in which they engage must have a clear relationship to the objectives and tests. Furst (1958) showed this relationship graphically.



There are several purposes for classroom testing. Tests provide *summative data* for the instructor to show differences among student scores and indicate student performance against predetermined performance criteria. This can be shown as a grade or P/F. They provide *formative data* as feedback to the instructor for the purpose of improving instruction. Tests also provide *diagnostic data* for the student to indicate strengths and weaknesses in relation to mastering the learning objectives.

From the viewpoint of the student, testing serves several purposes. One is to provide feedback *during* a course to show where additional mastery is needed relative to the performance objectives. Another purpose is to motivate the student by demonstrating the degree to which the student has mastered the material.

NORM-REFERENCED TESTING

Test results can be reported as norm-referenced or as criterion-referenced.

Norm-referenced testing is the more common method. Based on the assumptions of classic parametric statistics, it uses a comparative approach in which one student's performance is compared to another in reporting data. It poses the question, "Did John do as well as Mary and Bob?" Student scores are reported at the 90th or some lower percentile compared to all other students. Norm-referenced tests are designed to maximize variance among student test scores. They promote the concept of average score distribution in which some students receive a high score, most a middle score, and a few a low score. In other words, this testing approach is designed to fail students or at least have a few score at a low level. It does not indicate what a student can, in fact, do. It only compares one student with others in the peer group. In this situation, test items are designed primarily to discriminate between high and low scores on any test. High achievement theoretically translates as "The pursuit of excellence." Low achievement indicates a lack of mastery or competence in a skill or knowledge base.

of my class (or peers) learned more or less than me." *Criterion-referenced testing* asks, "How much of the specified performance has the student learned?"

TESTING AS A MEANS TO IMPROVE INSTRUCTION

Instruction can be qualitatively improved through testing. If the learning performance objectives have been stated specifically before instruction began and if they are shared with the students participating in the instruction, *test results will provide the following data:*

1. Did the test item measure the learning performance objective?
2. To what degree did the student meet the criterion level in the objective?
3. Was the instructional activity adequate to assist the student to meet the criterion level?
4. Were the sequences of the instruction adequate?
5. Did the instruction provide enough examples?

TEST VALIDITY AND RELIABILITY

When developing a test, the instructor must deal with two closely related problems. The knowledge and skills to measure must be identified. This is called validity, which is the extent to which a test measures what it is supposed to measure.The second problem deals with constructing the most dependable means to measure the performance. This is called reliability, which is the degree to which performance, as shown in the test score, remains consistent when the test is taken repeatedly. A test cannot be valid unless it is also reliable. However, the reverse is not necessarily true. Test validity is a function of reliability. Test reliability is not a function of validity. Test validity addresses the question, "Are the items in a test specific to the task(s) they are expected to measure?" Four types of validity can be checked in relation to this question and its answer:

1. *Content Validity.* Which tasks or objectives do the test items represent and how close is this representation?
2. *Concurrent Validity.* With what other present observations of the students' performance do the tests scores agree and to what degree do they agree?
3. *Predictive Validity.* What kinds of future performance can be predicted by the test scores and how accurate is this prediction?
4. *Construct Validity.* This is the most theoretical type of validity and asks what other psychological features account for some degree of the test or performance.

Validity is not something "in the test." It is a bridge between learning and the quality of performance the test is expected to demonstrate.

PLANNING YOUR TEST

Before writing the actual test items, the instructor must develop a "blue print" for assigning the number of items to adequately test achievement of the learning performance objectives and the mastery of the course content.

The following chart is based on Bloom's Taxonomy (Bloom, 1956) and proposes a model for developing a test blueprint. Such a blueprint is the tool an instructor should use to build *balance*, or appropriate representation of all learning tasks into the test. Assigning the number items under the objectives to be achieved for

each unit of course content represents their relative weight or importance in evaluating a student's overall performance.

EXAMPLE OF A FINAL EXAM BLUEPRINT

Objective / Course Content	Define or State Concepts and Principles	Recognize Correct or Incorrect Applications of Concepts and Principles	Apply Principles to Solve a Given Problem	Analyze a Problem and Solve It	TOTAL ITEMS
1. Meeting the patient's needs	1				1
2. The nature of illness	2	1			3
3. Mental health and mental illness	2	1	4	3	10
4. The patient in surgery	3	3	3	5	14
5. Emergency and first aid	1	3	3	4	11
6. The older adult and dying patient	2	2	5	2	11
7. The critically ill and dying patient	4	2	6	6	18
8. Maternal health care	2	2	6	4	14
9. Child health care	2	2	3	5	12
10. Legal aspects of nursing	1	1	2	2	6
TOTAL ITEMS	20	17	32	31	100

MATCHING TEST ITEMS TO LEARNING PERFORMANCE OBJECTIVES

Performance objectives describe student performance in measurable language that specifies *what* is to be done; *how* it is to be done; and *when* it is satisfactorily done. Student assessment procedures (criterion tests) should be designed to measure the performance described in the performance objectives (Cyrs, 1987). These *criterion tests* are developed to reflect the circumstances of assessment described in the performance objective and allow us to judge when the student has performed satisfactorily or competently. The type of test will depend on the type of behavior described in the learning performance objective.

A procedure for writing criterion test items from learning performance objectives consists of ten steps:

1. Read the performance objectives carefully and note the specific behavior that the student is required to exhibit. Is further clarification necessary?
2. Is this behavior primarily cognitive, attitudinal, or manual?
3. If cognitive, does it require only memorization or a higher level of intellectual skill?
4. Is the objective content specific or content general?
5. Does the action call for the student to select an appropriate answer from several provided, as in multiple-choice, or does the student have to recall and write out the answer, as in essay or fill-in?

6. If the student is asked to remember and write out the answer, will a few words suffice or will the student have to write several paragraphs or a complete essay?
7. Are adequate criteria specified to judge the successful performance of this behavior?
8. Will the expected performance require interaction with other students or instructors?
9. Will the assessment of this objective require performance or will paper and pencil methods suffice?
10. Have all alternative assessment methods that would validly measure this objective been identified?

EXAMPLES OF PERFORMANCE OBJECTIVES DERIVED FROM STATEMENTS OF SUBJECT MATTER

Given statements of subject matter, it is possible to specify learning performance objectives at different intellectual skill levels.

PERFORMANCE OBJECTIVE	Explain the role of hormones vital to body functions.
INTELLECTUAL LEVEL	Comprehension. The student paraphrases the information provided.
SUBJECT MATTER	Insulin has three basic effects on carbohydrate metabolism.
PERFORMANCE OBJECTIVE	Describe how to administer prescribed medications to patients under hospital care.
INTELLECTUAL LEVEL	Comprehension. The student paraphrases information.
SUBJECT MATTER	Insulin prepared in a suspension should be rolled gently between the palms of the hands to mix, since shaking will form bubbles and result in inaccurate doses of the medication.
PERFORMANCE OBJECTIVE	Given specific treatments in medical-surgical situations, predict possible consequences.
INTELLECTUAL LEVEL	Application. The student must apply previous knowledge to new and unfamiliar situations.
SUBJECT MATTER	In treatment of pneumonia, prolonged use of oxygen at high concentration may cause a fibrosis of lung tissue, impairing the ability to expand.
PERFORMANCE OBJECTIVE	Given a disease, identify the symptoms.
INTELLECTUAL LEVEL	Comprehension. The student recalls information and paraphrases it. If the objective was, "Given the symptoms, identify the disease," students would work at the critical analysis level, evaluating and deciding which procedures and information to apply.
SUBJECT MATTER	Symptoms of hyperthyroidism include nervousness, apprehension, and irritability. The patient seems to be in constant motion–turning, twisting–and excessively concerned about things. The patient will have a rapid pulse even when at rest. Although the patient may have an increased appetite, there will be progressive weight loss due to the abnormal activity.

TYPES OF TESTS

There are two major classes of tests, which include performance tests and paper and pencil tests.

PERFORMANCE TESTS

Performance tests require the student to perform an actual task under simulated or real world conditions. The test utilizes rating or observation scales and assessment center methodology.

PAPER AND PENCIL TESTS: Essay Tests

Paper and pencil tests are either essay or objective. An essay test is occasionally called a discussion type test. The student responds to a question using his/her own words to express conclusions and the reasons for reaching them. During an objective test a statement is presented or a question is posed. The student must either select the correct response from a number of plausible alternative answers that are provided, or supply an answer limited to one word or short phrases.

Many instructors prefer the essay test because of its apparent ease of preparation relative to a 50- to 75-item multiple-choice test.

There are a number of advantages to an essay test. Many faculty feel that they are easy to prepare and administer. Essays support the development of expressive writing skills for the student as they prepare their answers. The essay can test a very wide range of cognitive skills, such as critical thinking, as well as allow the student to express feelings and emotions.

Countering these apparent advantages, there are a number of significant disadvantages. This test is very unreliable and subject to scoring bias. It is tedious and difficult to score accurately and it will require more time to grade. Articulate students usually score higher because of their excellent writing skills.

If the purpose of the essay is recall of factual information, it is not efficient or desirable. It is best used for the expression of critical thinking skills.

If an instructor decides to use essay questions, there are a number of guidelines that will be of benefit.

SUGGESTED GUIDELINES FOR THE ESSAY TEST

1. Select questions to sample a broad range of cognitive and affective objectives and content.
2. Require that all opinions be supported with data.
3. Be sure that students have well-developed writing skills.
4. Guide the student toward the desired response rather than using very general, open-ended questions.
5. Suggest the number of points the question is given and the recommend amount of time that should be spent.
6. Develop a *criterion checklist* against which the instructor will evaluate the test question:
   a. Cite how the answer should be organized;
   b. Determine how many points for writing skill as well as completeness of the answer;
   c. Determine key elements that should be covered in the answer and the number of points assigned to each element;
   d. Determine if spelling and grammar are important and how many points will be assigned;

e. Require the student to answer each question rather than provide a choice of questions since this will lessen reliability.

7. Problem-oriented and simulation questions lend themselves to this type of test.
8. Tuckman (1975) suggests that an essay test have a center structure:
   a. Statement of the situation,
   b. Statement of the problem,
   c. Response instructions.

BASIC SCORING TECHNIQUES

Scoring the essay test is always a tedious task but can be made simple and more reliable if:

1. Use an answer key or criterion check list.
2. Score the answers *question by question,* rather than student by student in a single sitting.
3. Conceal the identity of the student whose answer is being scored.
4. Writing style should not be confused with content.
5. Ask a colleague to rate a sample of the answers using the answer key.
6. Provide useful and constructive feedback on the answer.

TWO SCORING METHODS FOR EACH ESSAY ITEM

A. COMPONENT SCORING
   1. Using the criterion check list, identify significant points on each student's essay based on your ideal answer. *Assign a positive value.*
   2. Using the same criterion check list, identify inaccuracies and irrelevant points. *Assign a positive value.*
   3. Analyze the overall organizational structure of the answer and identify the student's ability to integrate the most significant points. *Assign a positive value.*
   4. Construct written comments to each student concerning the assessment of his/her response. *Assign a positive or negative value.*
   5. Total the positive and negative values for each item. *Record student score.*

B. COMPARATIVE QUALITY SCORING
   1. Quickly read each student answer.
   2. Sort answers into piles representing high, middle, and low quality.
   3. Carefully review each set of answers in their respective piles.
   4. Using an answering key or criterion check list, shift deserving answers into a more appropriate pile.
   5. Score each pile from highest to lowest quality.

PAPER AND PENCIL: Multiple-Choice Tests

One of the most popular and frequently used forms of paper and pencil test formats is the multiple-choice form. Although it appears fairly easy to develop, this is very deceiving. Multiple-choice tests are difficult and time-consuming if they deal with

intellectual skills above simple recall and comprehension. There are several advantages to a multiple-choice test. Students can be tested on a large sample of the course content in a relatively short period of time. They are quick and efficient to score by hand using an answer key but machine processing is usually available in the computer center. The test items can measure a wide range of intellectual skills from recall of factual data, understanding, application, and critical thinking. Guessing is reduced to one in five as opposed to a 50/50 chance in true/false items.

A disadvantage of multiple-choice tests is the difficulty of constructing them with good distractors. There is too often a great deal of ambiguity in the choice of the correct answer.

All multiple-choice items have three main parts. These include the stem (statement or question), which begins the multiple-choice item; the distractors (incorrect responses which usually number three); and the last part is the correct response.

GUIDELINES

If you choose to construct multiple-choice test items, there are a number of guidelines that will be beneficial to you. Grammatical form should be consistent in each of the distractors and the correct answer. Inconsistency could provide unnecessary cues to the students. The length of the responses must be consistent. Match each test item to the performance objective it is designed to evaluate in order to maintain validity. Write reasonable distractors that are within the comprehension of the students. Try to avoid unnecessary jargon unless it was covered during class. Avoid obvious clues in the stem which are repeated in the distractors. Use either four or five distractors and avoid the use of negative statements in the stem or response. Limit the use of the "keyed response" such as "all of the above" or "none of the above."

SAMPLES OF DIFFERENT TYPES OF MULTIPLE-CHOICE ITEMS

DESCRIPTIVE vs. LABEL RESPONSES

The term distal means

A. Nearest the origin of a structure
B. Farthest from the origin of a structure
C. Nearest the midline of a structure
D. Farthest from the midline of a structure

An increase in the overall size of a tissue or organ is

A. Atrophy
B. Extrophy
C. Hypertrophy
D. Dystrophy

NEGATIVE STEM

It is not a function of the cerebellum to

A. Regulate body temperature
B. Maintain body balance
C. Coordinate working of muscles
D. Aid in maintaining muscle tone

BEST ANSWER

A reflex is best described as

A. The response one gets from tapping the kneecap
B. A sign of meningeal irritation
C. An involuntary muscular contraction in response to a stimulus
D. A voluntary movement of the skeletal muscle

INTRODUCTORY SENTENCE

A patient is admitted for an exploratory laparotomy. The surgical preparation would include shaving

A. From the nipple line to the perineum
B. The perineal area
C. From the umbilicus to the perineum
D. From the scapular area to the perineum

REPETITIOUS

Which is the best definition of a vein?

A. A blood vessel carrying blood going to the heart
B. A blood vessel carrying blue blood
C. A blood vessel caring impure blood
D. A blood vessel carrying blood away from the heart

CASE STUDY APPROACH

Ms. Andrews, age 21, is admitted to the hospital with a history of general malaise, nausea, vomiting, and evidence of jaundice. The admitting diagnosis is infectious hepatitis.

Hepatitis is an inflammation of the

A. Gall bladder
B. Liver
C. Small intestine
D. Stomach

PAPER/PENCIL TESTS: True/False Items

The true/false test item is somewhat similar to a multiple-choice item. The true/false item is a two-choice test item.

The advantages of the true/false test item includes the ease of construction as well as the ease of scoring by hand, with an answer key, or by computer. This type of test item allows for the fastest student response and can cover a broad range of subject matter. True/false items can form the basis for a future pool of multiple-choice items.

An obvious disadvantage is the susceptibility to guessing with a 50% chance for a correct response. Unfortunately, the ease of preparation leads to an overabundance of low level recall items that are often very ambiguous. If used exclusively they will overly influence students to learn lower level skills.

Well-constructed true/false test items can measure knowledge at a variety of levels.

COGNITIVE LEVEL		SAMPLE TRUE/FALSE TEST ITEMS
FACT	(T)	Hodgkin's disease is characterized by an increase in the size of lymph nodes.
EXPLANATION	(F)	A paralyzed patient must be watched very carefully for evidence of pressure sores, since nourishment to the area is decreased.
COMPUTATION	(T)	In order to convert a Fahrenheit reading to a centigrade reading you would subtract 32 from the Fahrenheit reading and multiply the fraction by 5/9.
APPLICATION	(F)	In taking blood pressure, you must take the systolic reading at the point where the sound stops.
ANALYSIS	(T)	Signs of increased intercranial pressure are indicated by vomiting, elevated blood pressure, slowing of pulse, slowing of respirations, and unequal pupils.
JUDGMENT	(F)	It is better for the nurse to have children of her own if she expects to have an adequate understanding of the sick child.

THE ARGUMENT FOR TRUE/FALSE TESTS

The basis for educational achievement is the command of useful verbal knowledge. All verbal knowledge can be expressed in true or false verbal propositions or sentences. The extent of a person's command of a particular area of knowledge is indicated by his/her success in determining the truth or falseness of related propositions. Having command of knowledge means one can use it to make decisions, draw logical inferences, or solve problems. It is knowledge that is available for use. True/false tests reflect usable knowledge.

HOW TO PREPARE EFFECTIVE ITEMS

1. Locate sources presenting bits of special knowledge: paragraphs, written procedures, etc.

   Paragraph:
   "One of the effects of digitalis is to make the heart beat slower. Therefore, the nurse must always check the pulse and if the rate is below 60, should withhold the medication until she notifies the physician."

2. Derive a proposition on which to base test items.

   Proposition:
   Digitalis must not be administered by the nurse to the patient if the pulse rate is below 60.

3. Restate the original idea in different words: the true and false versions are written in pairs. The false item is a contradiction that is worded to sound plausible.

   Items:
   (T) In giving digitalis the nurse must always take and record the pulse.
   (F) In giving digitalis the nurse must always take and record the blood pressure.

4. Interpret the basic idea in terms of another one similar to the original.

More items:
(T) A patient is given digitalis to slow the heart rate.
(F) A patient is given digitalis to increase the heart rate.

OR

5. Develop another way of looking at the basic idea: true and false version. If possible, create novel situations to exemplify the proposition.

More items:
(T) If a patient does not exhibit signs of bradycardia, the nurse will administer the prescribed amount of digitalis.
(F) The nurse will administer the prescribed amount of digitalis only when the patient exhibits signs of bradycardia.

CHARACTERISTICS OF GOOD TRUE/FALSE ITEM

Good true/false test items are based on an important idea rather than trivia. They test more understanding than rote memory. The correct answer is defensible and not ambiguous. The wrong answer would seem reasonable to someone who does not possess the appropriate knowledge. The test item is based on a single idea that is concise and clearly expressed.

CHECKPOINTS TO OBSERVE WHEN WRITING A TRUE/FALSE TEST ITEM

- Use more false statements (perhaps two false to one true).
- Minimize use of inadvertent clues that may be used by the test-wise student.
- Use phrases in false statements that give them the impression of truth.
- Write the content or statement in both true and false version.
- Avoid vague terms.
- Avoid qualified statements such as "often," "occasionally," "never," "may," or "generally."
- Make all items approximately the same length to avoid giving student cues that one type may be longer or shorter.
- Avoid tricky questions.
- True/false items are dependent on a variety of sentence forms.
- Testing at an appropriate level of achievement, in large part, is related to the sentence form.

The following chart may be used as a guide in selecting some of the key introductory words or phrases to be used in items testing various aspects of learning. Obviously, there are slight variations possible with words or phrases, as well as others appropriate to a particular aspect of learning. Since all objective items rely on command of knowledge that is essentially verbal, the chart is useful for other types of tests as well.

PERFORMANCE TASK	INTRODUCTORY WORDS OR PHRASES
	Any declarative statement.
RECALLING FACTS	All... Most... Many...
GENERALIZING	The difference between... Both...
COMPARING	If... When...
INFERRING	The larger... The higher... The lower... Making...
RELATING	is likely to... Increasing... tends to... How much... depends on...
EXPLAINING	The main reason... is to... The purpose of... is to...
EXEMPLIFYING	An example of...
ANALYZING	Observing... reveals that... Studying... indicates that...
PREDICTING	One could expect that...
APPLYING	To... one must... In order to... one must use... One method of... is to... One essential step... is to... The first step...
COMPUTING	Item includes numbers and requires computation or estimation.
EVALUATING	A good... It is better to... than... The best proportion... The maximum variability... The easiest method... is to... While easy to... it is not... It is difficult to... It is possible... It is reasonable...

PAPER AND PENCIL TESTS: Completion or Short Answer Items

The completion item requires the student to fill in a short answer with his/her own words. Described as a constructed response test, it is a derivative of the multiple-choice test item.

Stem: The parasympathetic nervous system is part of the_____

Answer: autonomic nervous system.

The stem can be presented as an incomplete sentence as in the above example or in the form of a question:

Stem: Which part of the brain controls body temperature?_____

EXAMPLES

CALLED FILL-IN OR SHORT ANSWER ITEMS

Stem: Blood pumped from the left ventricle flows through the _____.

Answer: aortic valve.

GIVEN THE STEM, PARTIAL STATEMENTS, QUESTIONS, OR TERMS

Stem: If a patient is unable to lift his hips to get onto the bed pan, what should the nurse do?

Answer: Roll the patient on his side, put the bed pan in place, and roll the patient back on it.

THE STUDENT SUPPLIES A TERM, PHRASE, NUMBER, OR OTHER SYMBOL

Stem: The peak effect of insulin is reached in _____ hours.

ANY COMBINATION OF STEM AND ANSWER.
The student in each case is required to recall the response that the instructor has predetermined as correct. These test items should always be preceded with specific directions:

EXAMPLE

On the blank following each of the questions, partial statements, or words, you are to write the word, short phrase, or number that seems most appropriate.

If the answer is too obviously cued in a multiple-choice (or selected response) test item, then a completion or short answer item may be preferred.

Compare the two previous examples of completion items with their multiple-choice versions below.

The parasympathetic nervous system is part of the
A. central nervous system
B. autonomic nervous system
C. sympathetic nervous system
D. all of the above

Body temperature is controlled by
A. thalamus
B. thyroid
C. hypothalamus
D. none of the above

ADVANTAGES

- Easy to write certain items.
- Student must recall the answer (more apparent than real)

DISADVANTAGES

- Limited to questions that can be answered with a word, phrase, or symbol.
- Scoring tends to be tedious and subjective.
- Difficult to write items requiring short, sometimes specific answers that test higher learning levels.

GUIDELINES

1. Allow sufficient space for the student response.
2. Keep all response blanks of equal length to avoid cuing.
3. A question format is often more desirable than a statement completion. The grammatical style of the latter could influence the choice of answer.

PAPER AND PENCIL TESTS: The Matching Item

Matching items can measure a range of behavior but are most commonly used to measure recall behavior. A matching test consists of a set of "stems" or "questions" on the left hand side to which a set of responses on the right hand side are matched by the student.

ADVANTAGES

- Fairly easy to prepare.
- Efficient in the respect that the same set of responses can be used with several similar "stems."

DISADVANTAGES

- Difficult to measure higher levels of learning.
- Usually too many tricky questions.

GUIDELINES FOR EFFECTIVE MATCHING ITEM CONSTRUCTION

1. Keep the style and content of the stem and response columns homogeneous. It is preferable to use short responses of even length to avoid unnecessary clues.
2. Limit the number of stems from six to ten. Additional matches introduce fatigue and confusion to the matching process.

EXAMPLES OF MATCHING TEST ITEMS

Match the developmental phase with the appropriate chronological period of a person's life:

1. Oedipal	_____ a. 0-1 year
2. Maturity	_____ b. 1-3 years
3. Oral	_____ c. 3-6 years
4. Adolescence	_____ d. 6-12 years
5. Latency	_____ e. 12-21 years
6. Anal	_____ f. 21 years
	_____ g. 35 years

For each term in Column A (premises), select the statement in Column B (responses that best define the terms in Column A). Mark your answer on the appropriate line in Column A. There is only one correct definition for each term.

A	B
_____ 1. Split-half reliability.	a. An estimate of the degree of correlation between alternate forms of a test.
_____ 2. Coefficient of stability.	b. An estimate of the relationship between two measures of the same person.
_____ 3. Coefficient of equivalence.	c. A measure of the internal consistency of test results.
_____ 4. Concurrent validity.	d. An estimate of the correlation between the results of two different measures obtained at the same time.
_____ 5. Predictive validity.	e. An estimate of the correlation between the results of some measure and the results of some criterion of measure obtained at a later date.

ANSWERS TO THE EXAMPLES ABOVE
1. c 2. g 3. a 4. d 5. e

In the following items you are to designate the time period during which a particular sign or symptom of pregnancy can best be used. For each item, place an "X" in the answer space.

A. 1-3 months
B. 4-6 months
C. 7-10 months

	A	B	C
17. Colostrum expressed	—	—	—
18. Male	—	—	—
19. Goodell's sign	—	—	—
20. Braxton Hicks contractions	—	—	—
21. Cervical mucus positive fern	—	—	—
22. Ballottement	—	—	—
23. Leaking colostrum	—	—	—

Directions: On the blank before the name of each disease place the letter that precedes the medication for which the disease is known to respond best.

Disease	Medication
_____ 1. Parkinson'	a. atropine
_____ 2. Addison'	b. levodope
_____ 3. Multiple Sclerosis	c. corticoid preparation
	d. unknown

FINAL TESTING CONSIDERATIONS

Regardless of the type or form of test that you have chosen, each item should be based on a stated learning performance objective to improve the test validity. In

constructing a test the instructor should make every effort to reduce test error. Toward this end several guidelines are suggested:

1. Advise the students in advance which test format will be used.
2. Before the test inform the students which learning performance objectives and content areas will be covered.
3. Review the grading policy and value of each test as it relates to the final grade.
4. Arrange several of the easiest items at the beginning of the test to reduce anxiety.
5. Arrange all test items in random order.
6. Make test instructions specific and explicit.
7. If possible, have a colleague check the test items for ambiguity.

ALTERNATIVE STUDENT ASSESSMENT PROCEDURES

Paper and pencil testing has traditionally taken the forms of the test types just described. From the perspective of teaching faculty, these appear to be safest. However, there are a number of other student assessment techniques that deserve consideration.

Open Book Examinations • This type of assessment procedure allows the student to use any references available, including the textbook, class notes, and handouts. Rather than memorizing a lot of material, the instructor wants the students to apply skills by using formulas, tables, graphs, and so forth. The emphasis is on application of essential skills rather than memory and restatement of data. This technique tends to reduce student anxiety. The research indicates that there is no clear benefit in achievement of learning outcomes (Boniface 1985).

One variation of the open book test is to allow the student to use crib sheets or provide them with the essential formulas, tables, or graphs, without the benefit of the open book.

Some instructors combine the open book with the standard classroom test.

Take-Home Examinations • These allow the student to take the examination home to complete within a prescribed time period. The student is expected to synthesize a lot of background reading, use references, and produce a logical summary to the assignment. This type of exam is useful when there is a great deal of writing required on the part of the student. Using this format, the student can explore much broader implications of an issue, since more time is available for reflection and bridge-building of ideas. To be fair to all students, the instructor should specify the amount of time that should be spent on each answer and the number of pages expected. Grading on this type of examination must be based on specific criteria that are shared with the students before they take the examination. The take-home examination can be combined with the in-class test.

Oral Examinations • Usually used at the graduate level, they are rarely used for undergraduate testing. To be absolutely objective, there should be

more than one instructor administering an oral exam. They are time-consuming to give and difficult to grade. The oral exam creates undue stress and anxiety on the part of the students, who must "think on their feet." Most students have not been trained on how to take an oral examination and are easily intimidated, since it is very difficult to bluff or guess. If this exam format is used with a number of students, the instructor should use a prepared list of questions that are asked of all the students. Students should be informed in advance of the test exactly what is expected and any time parameters under which they must respond.

Mastery Testing • This approach allows the student to repeat a different form of the test a number of times. It is widely used in the teaching strategy known as the Keller Plan or Personalized System of Instruction. A large bank of test items is required for the alternative tests, which must be equivalent in difficulty level. Most students are required to retake the test within three to five days of the original testing.

Collaborative Testing • A small group of students, usually three to five, consult with each other during the test. Although they question and teach each other, they must turn in individual answer sheets. This type of participatory testing still requires individual accountability on the part of each student. Questions must be formatted at the critical thinking or application level of thinking. Knowledge and simple comprehension questions are inappropriate since these have only one correct response.

Paired Testing • This is a form of collaborative testing using pairs of students.

Performance Testing • Students must not only know and understand a procedure, they must apply it in front of the instructor, who assesses the processes used by the students as well as the final product. The instructor must have available a set of prespecified performance criteria that matches the learning performance objectives. These are shared with and explained to the student prior to assessment time. The performance criteria are usually in the form of lists or rating scales. Lists require the instructor to determine a yes-no check as to the presence or absence of a skill. Rating scales, on the other hand, allow the instructor to specify the degree of accuracy on a continuous scale.

Journal • This is a cross between a student notebook and a writer's diary (Stanley 1991). The students record their reactions to reading assignments as well as class sessions. They identify ideas and connections among ideas that they have trouble understanding. They list questions they have and speculations about the material. The students are required to reflect on and think about the material they are studying. The journal can be kept outside of class or the instructor can provide the last five minutes of each class period for the students

to make journal entries. Journals are collected periodically and returned with comments from the instructor.

The instructor needs to provide specific criteria that will be used to judge the journal entries.

Some instructors have provided evaluative criteria for students or had the students generate their own criteria for peer review. In some instances, the instructors have had the students grade their own journals.

Portfolio • A student portfolio is defined as "a purposeful collection of student work that tells the story of the student's efforts, progress, or achievement in a given area" (Arter and Spandel 1992). The portfolio, kept by the student, provides a sample of the student's work and progress toward the learning performance objectives. The portfolio provides a sample of the best work of the student. It can contain poetry, papers, artifacts, videos of performance in theatre or music, speeches, artwork, musical compositions, and so forth. The portfolio is collected at mid-term and again at the completion of the course for evaluation and comment by the instructor. Portfolios should always be returned to the student.