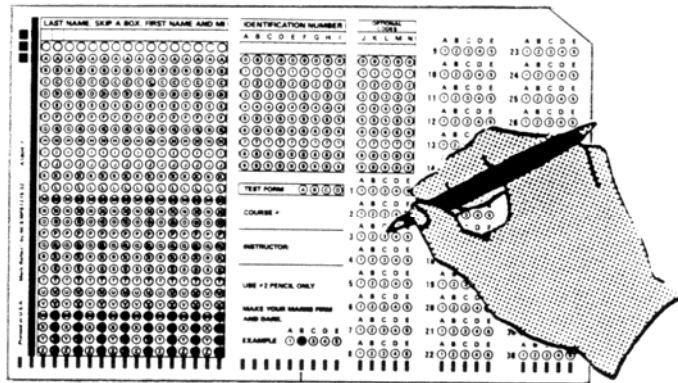


Improving Your Test Questions

EES Memorandum #27



Prepared by **John C. Ory**
Office of Instructional Resources
University of Illinois
Urbana-Champaign

Revised and Distributed by
Evaluation and Examination Service
The University of Iowa
(319) 335-0356

IMPROVING YOUR TEST QUESTIONS

Choosing the appropriate type of test item to measure students' understanding of course material and their achievement of course goals can often be as difficult a task as writing the items themselves. The purpose of this Bulletin is to provide information about the uses, advantages and limitations of the various items types and to help you develop specific skills in writing each type of item.

This bulletin is divided into the following sections:

I. Choosing Between Objective and Subjective Test Items.....	2
II. Suggestions for Using and Writing Test Items	
Multiple-Choice	6
True-False	11
Matching	14
Completion.....	17
Essay	19
Problem Solving.....	23
Performance	26
III Checklist for Evaluating Test Items.....	28
IV. References for Further Reading.....	31

I. CHOOSING BETWEEN OBJECTIVE AND SUBJECTIVE TEST ITEMS.

There are two general categories of test items: (a) objective items which require students to select the correct response from several alternatives or to supply a word or short phrase to answer a question or complete a statement; and (2) subjective or essay items which permit the student to organize and present an original answer. Objective items include multiple-choice, true-false, matching and completion, while subjective items include short-answer essay, extended-response essay, problem solving and performance test items. For some instructional purposes one or the other item types may prove more efficient and appropriate. To begin our discussion of the relative merits of each type of test item, test your knowledge of these two item types by answering the following questions.

Test Item Quiz

(Circle the correct answer)

1. Essay exams are easier to construct than are objective exams. T F ?
2. Essay exams require more thorough student preparation and student time than objective exams. T F ?
3. Essay exams require writing skills where objective exams do not. T F ?
4. Essay exams teach a person how to write. T F ?
5. Essay exams are more subjective in nature than are objective exams. T F ?
6. Objective exams encourage guessing more so than essay exams. T F ?
7. Essay exams limit the extent of content covered. T F ?
8. Essay and objective exams can be used to measure the same content or ability. T F ?
9. Essay and objective exams are both good ways to evaluate a student's level of knowledge. T F ?

Quiz Answers

1. TRUE - Essay items are generally easier and less time consuming to construct than are most objective test items. Technically correct and content appropriate multiple-choice and true-false test items require an extensive amount of time to write and revise. For example, a professional item writer produces only 9 - 10 good multiple-choice items in a day's time.
2. ? - According to research findings it is still undetermined whether or not essay tests require or facilitate more thorough (or even different) student study preparation.
3. TRUE - Writing skills do affect a student's ability to communicate the correct "factual" information through an essay response. Consequently, students with good writing skills have an advantage over students who have difficulty expressing themselves through writing.
4. FALSE - Essays do not teach a student how to write but they can emphasize the importance of being able to communicate through writing. Constant use of essay tests may encourage the knowledgeable student with poor writing skills to improve his/her writing ability in order to improve performance.
5. TRUE - Essays are more subjective in nature due to their susceptibility to scoring influences. Different readers can rate identical responses differently, the same reader can rate the same paper differently over time, the handwriting, neatness or punctuation can unintentionally affect a paper's grade and the lack of anonymity can affect the grading process. While impossible to eliminate, scoring influences or biases can be minimized through procedures discussed later in this bulletin.
6. ? - Both item types encourage some form of guessing. Multiple-choice, true-false and matching items can be correctly answered through blind guessing, yet essay items can be responded to satisfactorily through well written bluffing.
7. TRUE - Due to the extent of time required by the student to respond to an essay question, only a few essay questions can be included on a classroom exam. Fewer items on a test limits content coverage.
8. TRUE - Both item types can measure similar content or learning objectives. Research has shown that students respond almost identically to essay and objective test items covering the same content. Studies by Sax & Collet (1968)¹ and Paterson (1926)² conducted forty-two years apart reached the same conclusion:

¹ Gilbert Sax and LeVerne S. Collet, "An Empirical Comparison of the Effects of Recall and Multiple-Choice Tests on Student Achievement," *Journal of Educational Measurement*, vol.5 (1968) 169-73.

² Donald G. Paterson, "Do New and Old Type Examinations Measure Different Mental Functions?" *School and Society*, vol. 24. (August, 21, 1926), 246-48.

“...there seems to be no escape from the conclusions that the two types of exams are measuring identical things.” (Paterson, p.246)

This conclusion should not be surprising; after all, a well written essay item requires that the student (1) have a store of knowledge, (2) be able to relate facts and principles, and (3) be able to organize such information into a coherent and logical written expression, whereas an objective test item requires that the student (1) have a store of knowledge, (2) be able to relate facts and principles, and (3) be able to organize such information into a coherent and logical choice among several alternatives.

9. TRUE - Both objective and essay test items are good devices for measuring student achievement. However, as seen in the previous quiz answer, there are particular measurement situations where one item type is more appropriate than the other. Following is a set of recommendations for using either objective or essay test items: (Adapted from Robert L. Ebel and David A. Frisbie, *Essentials of Educational Measurement*, 1991, p. 123).

WHEN TO USE ESSAY OR OBJECTIVE TESTS

Essay tests are especially appropriate when:

- ? the group to be tested is small and the test is not to be reused.
- ? you wish to encourage and reward the development of student skill in writing.
- ? you are more interested in exploring the student’s attitudes than in measuring his/her achievement.
- ? you are more confident of your ability as a critical and fair reader than as an imaginative writer of good objective test items.
- ? there is more pressure for speedy test preparation than for speedy reporting of scores.

Either essay or objective tests can be used to:

- ? measure almost any important educational achievement that any paper and pencil test can measure.
- ? encourage students to study for understanding of principles, organization and integration of ideas, and application of knowledge to the solution of problems.
- ? test ability to think critically.
- ? test ability to select relevant facts and principles and to integrate them toward the solution of complex problems.

In addition to the preceding suggestions, it is important to realize that certain item types are better suited than others for measuring particular learning objectives. For example, learning objectives requiring the student to demonstrate or to show, may be better measured by performance test items, whereas objectives requiring the student to explain or to describe may be better measured by essay test items. The matching of learning objective expectations with certain item types can help you select an appropriate kind of test item for your classroom exam as well as provide a higher degree of test validity (i.e., testing what is supposed to be tested). To further illustrate, several sample learning objectives and appropriate test items are provided.

Learning Objective

Most Suitable Test Item

The student will be able to categorize and name the parts of the human skeletal system.

Objective Test Item
(M-C, T-F, Matching)

The student will be able to critique and appraise another student's English composition on the basis of its organization.

Essay Test Item
(Extended-Response)

The student will demonstrate safe laboratory skills.

Performance Test Item

The student will be able to cite four examples of satire that Twain uses in Huckleberry Finn.

Essay Test Item
(Short-Answer)

After you have decided to use either an objective, essay or both objective and essay exam, the next step is to select the kind(s) of objective or essay items that you wish to include on the exam. To help you make such a choice, the different kinds of objective and essay items are presented in the following section of this bulletin. The various kinds of items are briefly described and compared to one another in terms of their advantages and limitations for use. Also presented is a set of general suggestions for the construction of each item variation.

II. MULTIPLE-CHOICE TEST ITEMS

The multiple-choice item consists of two parts: (a) the stem, which identifies the question or problem and (b) the response alternatives. Students are asked to select the one alternative that best completes the statement or answers the question. For example,

Sample Multiple-Choice Item

(a) Item Stem: Which of the following is a chemical change?

- (b) Response Alternatives:
- a. Evaporation of alcohol
 - b. Freezing of water
 - *c. Burning of oil
 - d. Melting of wax

Advantages in Using Multiple-Choice Items

Multiple-choice items can provide...

- ? versatility in measuring all levels of cognitive ability.
- ? highly reliable test scores.
- ? scoring efficiency and accuracy.
- ? objective measurement of student achievement or ability.
- ? a wide sampling of content or objectives.
- ? a reduced guessing factor when compared to true-false items.
- ? different response alternatives which can provide diagnostic feedback.

Limitations in Using Multiple-Choice Items

Multiple-choice items ...

- ? are difficult and time consuming to construct.
- ? lead an instructor to favor simple recall of facts.
- ? place a high degree of dependence on the student's reading ability and instructor's writing ability.

SUGGESTIONS FOR WRITING MULTIPLE-CHOICE TEST ITEMS

The Stem

1. When possible, state the stem as a direct question rather than as an incomplete statement.

Undesirable: *Alloys are ordinarily produced by...*

Desirable: *How are alloys ordinarily produced?*

2. Present a definite, explicit and singular question or problem in the stem.

Undesirable: *Psychology...*

Desirable: *What is the science of mind and behavior called?*

3. Eliminate excessive verbiage or irrelevant information from the stem.

Undesirable: *While ironing her formal, Jane burned her hand accidentally on the hot iron. This was due to a transfer of heat by...*

Desirable: *Which of the following ways of heat transfer explains why Jane's hand was burned after she touched a hot iron?*

4. Include in the stem any word(s) that might otherwise be repeated in each alternative.

Undesirable: *In national elections in the United States the President is officially _____.
a. chosen by the people
b. chosen by members of Congress
c. chosen by the House of Representatives
d. chosen by the Electoral College

Desirable: *In national elections in the United States the President is officially chosen by _____.
a. the people
b. members of Congress
c. the House of Representatives
d. the Electoral College

5. Use negatively stated stems sparingly. When used, underline and/or capitalize the negative word.

Undesirable: *Which of the following is not cited as an accomplishment of the Kennedy administration?*

Desirable: *Which of the following is **NOT** cited as an accomplishment of the Kennedy administration?*

Item Alternatives

6. Make all alternatives plausible and attractive to the less knowledgeable or skillful student.

What process is most nearly the opposite of photosynthesis?

Undesirable

- a. Digestion
- b. Relaxation
- *c. Respiration
- d. Exertion

Desirable

- a. Digestion
- b. Assimilation
- *c. Respiration
- d. Catabolism

7. Make the alternatives grammatically parallel with each other, and consistent with the stem.

Undesirable: *What would do most to advance the application of atomic discoveries to medicine?*

- *a. Standardized techniques for treatment of patients.
- b. Train the average doctor to apply radioactive treatments.
- c. Remove the restrictions on the use of radioactive substances.
- d. Establishing hospitals staffed by highly trained radioactive therapy specialists.

Desirable: *What would do the most to advance the application of atomic discoveries to medicine?*

- *a. Development of standardized techniques for treatment of patients.
- b. Training of the average doctor in application of radioactive treatments.
- c. Removal of restrictions on the use of radioactive substances.
- d. Addition of trained radioactive therapy specialists to hospital staffs.

8. Make the alternatives mutually exclusive.

Undesirable: *The daily minimum required amount of milk that a 10 year old child should drink is _____.*

- a. 1 - 2 glasses
- *b. 2 - 3 glasses
- c. 3 - 4 glasses
- d. at least 4 glasses

Desirable: *What is the daily minimum required amount of milk a 10 year old child should drink?*

- a. 1 glass
- b. 2 glasses
- *c. 3 glasses
- d. 4 glasses

9. When possible, present alternatives in some logical order (e.g., chronological, most to least, alphabetical).

At 7 a.m. two trucks leave a diner and travel north. One truck averages 42 miles per hour and the other truck averages 38 miles per hour. At what time will they be 24 miles apart?

Undesirable

- a. 6 p.m.
- b. 9 p.m.
- c. 1 a.m.
- *d. 1 p.m.
- e. 6 a.m.

Desirable

- a. 1 a.m.
- b. 6 a.m.
- c. 9 a.m.
- *d. 1 p.m.
- e. 6 p.m.

10. Be sure there is only one correct or best response to the item.

Undesirable: *The two most desired characteristics in a classroom test are validity and _____.*

- a. precision
- *b. reliability
- c. objectivity
- *d. consistency

Desirable: *The two most desired characteristics in a classroom test are validity and _____.*

- a. precision
- *b. reliability
- c. objectivity
- d. standardization

11. Make alternatives approximately equal in length.

Undesirable: *The most general cause of low individual incomes in the United States is _____.*

- *a. the lack of valuable productive services to sell
- b. unwillingness to work
- c. automation
- d. inflation

Desirable: *What is the most general cause of low individual incomes in the United States?*

- *a. The lack of valuable productive services to sell.
- b. The population's overall unwillingness to work.
- c. The nation's increased reliance on automation.
- d. An increasing national level of inflation.

12. Avoid irrelevant clues such as grammatical structure, well known verbal associations or connections between stem and answer.

Undesirable: *A chain of islands is called an _____.*
(grammatical clue) **a. archipelago*
b. peninsula
c. continent
d. isthmus

Undesirable: *The reliability of a test can be estimated by a coefficient of _____.*
(verbal association clue) *a. measurement*
**b. correlation*
c. testing
d. error

Undesirable: *The height to which a water dam is built depends on the _____.*
(connection between stem and answer clue) *a. length of the reservoir behind the dam*
b. volume of water behind the dam
**c. height of water behind the dam*
d. strength of the reinforcing wall

13. Use at least four alternatives for each item to lower the probability of getting the item correct by guessing.
14. Randomly distribute the correct response among the alternative positions throughout the test having approximately the same proportion of alternatives a, b, c, d and e as the correct response.
15. Use the alternatives “none of the above” and “all of the above” sparingly. When used, such alternatives should occasionally be used as the correct response.

SUGGESTIONS FOR WRITING TRUE-FALSE TEST ITEMS

1. Base true-false items upon statements that are absolutely true or false, without qualifications or exceptions.

Undesirable: *Nearsightedness is hereditary in origin.*

Desirable: *Geneticists and eye specialists believe that the predisposition to nearsightedness is hereditary.*

2. Express the item statement as simply and as clearly as possible.

Undesirable: *When you see a highway with a marker that reads, "Interstate 80" you know that the construction and upkeep of that road is maintained by the state and federal government.*

Desirable: *The construction and maintenance of interstate highways is provided by both state and federal governments.*

3. Express a single idea in each test item.

Undesirable: *Water will boil at a higher temperature if the atmospheric pressure on its surface is increased and more heat is applied to the container.*

Desirable: *Water will boil at a higher temperature if the atmospheric pressure on its surface is increased.*

and/or

Water will boil at a higher temperature if more heat is applied to the container.

4. Include enough background information and qualifications so that the ability to respond correctly to the item does not depend on some special, uncommon knowledge.

Undesirable: *The second principle of education is that the individual gathers knowledge.*

Desirable: *According to John Dewey, the second principle of education is that the individual gathers knowledge.*

5. Avoid lifting statements from the text, lecture or other materials so that memory alone will permit a correct answer.

Undesirable: *For every action there is an opposite and equal reaction.*

Desirable: *If you were to stand in a canoe and throw a life jacket forward to another canoe, chances are your canoe would jerk backward.*

6. Avoid using negatively stated item statements.

Undesirable: *The Supreme Court is NOT composed of nine justices.*

Desirable: *The Supreme Court is composed of nine justices.*

7. Avoid the use of unfamiliar vocabulary.

Undesirable: *According to some politicians, the raison d'être for capital punishment is retribution.*

Desirable: *According to some politicians, justification for the existence of capital punishment is retribution.*

8. Avoid the use of specific determiners which would permit a test-wise but unprepared examinee to respond correctly. Specific determiners refer to sweeping terms like “all,” “always,” “none,” “never,” “impossible,” “inevitable,” etc. Statements including such terms are likely to be false. On the other hand, statements using qualifying determiners such as “usually,” “sometimes,” “often,” etc., are likely to be true. When statements do require the use of specific determiners, make sure they appear in both true and false items.

Undesirable: *All sessions of Congress are called by the President (F)*

The Supreme Court is frequently required to rule on the constitutionality of a law. (T)

An objective test is generally easier to score than an essay test. (T)

Desirable: (When specific determiners are used reverse the expected outcomes.)

The sum of the angles of a triangle is always 180 degrees. (T)

Each molecule of a given compound is chemically the same as every other molecule of that compound. (T)

The galvanometer is the instrument usually used for the metering of electrical energy used in a home. (F)

9. False items tend to discriminate more highly than true items. Therefore, use more false items than true items (but no more than 15% additional false items).

MATCHING TEST ITEMS

In general, matching items consist of a column of *stimuli* presented on the left side of the exam page and a column of *responses* placed on the right side of the page. Students are required to match the response associated with a given stimulus. For example,

Sample Matching Test Item

Directions: On the line to the left of each factual statement, write the letter of the principle which best explains the statement's occurrence. Each principle may be used more than once.

<u>Factual Statements</u>	<u>Principles</u>
_____ 1. Fossils of primates first appear in the Cenozoic rock strata, while trilobite remains are found in Protorozoic rocks.	a. There have been profound changes in the climate on earth
_____ 2. The Arctic and Antarctic regions are sparsely populated.	b. Coordination and integration of action is generally slower in plants than in animals.
_____ 3. Plants have no nervous system.	c. There is an increasing complexity of structure and functions from lower to higher forms of life
_____ 4. Large coal beds exist in Alaska.	d. All life comes from life and produces its own kind of living organisms.
	e. Light is a limiting factor to life.

Advantages in Using Matching Items

Matching items...

- ? require short periods of reading and response time, allowing you to cover more content.
- ? provide objective measurement of student achievement or ability.
- ? provide highly reliable test scores.
- ? provide scoring efficiency and accuracy.

Limitations in Using Matching Items

Matching items...

- ? have difficulty measuring learning objectives requiring more than simple recall of information.
- ? are difficult to construct due to the problem of selecting a common set of stimuli and responses.

SUGGESTIONS FOR WRITING MATCHING TEST ITEMS

1. Include directions which clearly state the basis for matching the stimuli with the responses. Explain whether or not a response can be used more than once and indicate where to write the answer.

Undesirable: *Directions:* Match the following.

Desirable: *Directions:* On the line to the left of each identifying location and characteristics in Column I, write the letter of the country in Column II that is best defined. Each country in Column II may be used more than once.

2. Use only homogeneous material in matching items.

Undesirable: *Directions:* Match the following.

- | | |
|---|---------------------|
| 1. ____ Water | A. NaCl |
| 2. ____ Discovered Radium | B. Fermi |
| 3. ____ Salt | C. NH ₃ |
| 4. ____ Year of the 1st Nuclear Fission
by man | D. H ₂ O |
| 5. ____ Ammonia | E. 1942 |
| | F. Curie |

Desirable: *Directions:* On the line to the left of each compound in Column I, write the letter of the compound's formula presented in Column II. Use each formula only once.

- | Column I | Column II |
|-----------------------|-----------------------------------|
| 1. ____ Water | A. H ₂ SO ₄ |
| 2. ____ Salt | B. HCl |
| 3. ____ Ammonia | C. NaCl |
| 4. ____ Sulfuric Acid | D. H ₂ O |
| | E. H ₂ HCl |

3. Arrange the list of responses in some systematic order if possible (e.g., chronological, alphabetical).

Directions: On the line to the left of each definition in Column I, write the letter of the defense mechanism in Column II that is described. Use each defense mechanism only once.

<i>Column I</i>	<i>Undesirable</i>	<i>Desirable</i>
	<i>Column II</i>	
____ 1. <i>Hunting for reasons to support one's beliefs.</i>	a. <i>Rationalization</i>	a. <i>Denial of reality</i>
____ 2. <i>Accepting the values and norms of others as one's own even when they are contrary to previously held values.</i>	b. <i>Identification</i>	b. <i>Identification</i>
____ 3. <i>Attributing to others one's own unacceptable impulses, thoughts and desires.</i>	c. <i>Projection</i>	c. <i>Introjection</i>
____ 4. <i>Ignoring disagreeable situations, topics, sights.</i>	d. <i>Introjection</i>	d. <i>Projection</i>
	e. <i>Denial of Reality</i>	e. <i>Rationalization</i>

4. Avoid grammatical or other clues to the correct response.

Undesirable: Directions: Match the following in order to complete the sentences on the left.

____ 1. <i>Igneous rocks are formed</i>	A. <i>a hardness of 7.</i>
____ 2. <i>The formation of coal requires</i>	B. <i>with crystalline rock.</i>
____ 3. <i>A geode is filled</i>	C. <i>a metamorphic rock.</i>
____ 4. <i>Feldspar is classified as</i>	D. <i>heat and pressure.</i>
	E. <i>through the solidification of molten lava.</i>

5. Keep matching items brief, limiting the list of stimuli to under 10.
6. Include more responses than stimuli to help prevent answering through the process of elimination.
7. When possible, reduce the amount of reading time by including only short phrases or single words in the response list.

COMPLETION TEST ITEMS

The completion item requires the student to answer a question or to finish an incomplete statement by filling in a blank with the correct word or phrase. For example,

Sample Completion Item

According to Freud, personality is made up of three major systems, the _____, the _____ and the _____.

Advantages in Using Completion Items

Completion items can...

- ? provide a wide sampling of content.
- ? efficiently measure lower levels of cognitive ability.
- ? minimize guessing as compared to multiple-choice or true-false items.
- ? usually provide an objective measure of student achievement or ability.

Limitations in Using Completion Items

Completion items...

- ? are difficult to construct so that the desired response is clearly indicated.
- ? have difficulty measuring learning objectives requiring more than simple recall of information.
- ? can often include more irrelevant clues than do other item types.
- ? are more time consuming to score when compared to multiple-choice or true-false items.
- ? are more difficult to score since more than one answer may have to be considered correct if the item was not properly written.

SUGGESTIONS FOR WRITING COMPLETION TEST ITEMS

1. Omit only significant words from the statement.

Undesirable: *Every atom has a central (core) called a nucleus.*

Desirable: *Every atom has a central core called a(n) (nucleus).*

2. Do not omit so many words from the statement that the intended meaning is lost.

Undesirable: *The _____ were to Egypt as the _____ were to Persia and as _____ were to the early tribes of Israel.*

Desirable: *The Pharaohs were to Egypt as the _____ were to Persia and as _____ were to the early tribes of Israel.*

3. Avoid grammatical or other clues to the correct response.

Undesirable: *Most of the United States' libraries are organized according to the (Dewey) decimal system.*

Desirable: *Which organizational system is used by most of the United States' libraries? (Dewey decimal)*

4. Be sure there is only one correct response.

Undesirable: *Trees which shed their leaves annually are (seed-bearing, common).*

Desirable: *Trees which shed their leaves annually are called (deciduous).*

5. Make the blanks of equal length.

Undesirable: *In Greek mythology, Vulcan was the son of (Jupiter) and (Juno).*

Desirable: *In Greek mythology, Vulcan was the son of (Jupiter) and (Juno).*

6. When possible, delete words at the end of the statement after the student has been presented a clearly defined problem.

Undesirable: *(122.5) is the molecular weight of $KClO_3$.*

Desirable: *The molecular weight of $KClO_3$ is (122.5).*

7. Avoid lifting statements directly from the text, lecture or other sources.

8. Limit the required response to a single word or phrase.

ESSAY TEST ITEMS

The essay test is probably the most popular of all types of teacher-written tests. In general, a classroom essay test consists of a small number of questions to which the student is expected to demonstrate his/her ability to (a) recall factual knowledge, (b) organize this knowledge and (c) present the knowledge in a logical, integrated answer to the question. An essay test item can be classified as either an extended-response or a short-answer essay item. The latter calls for a more restricted or limited answer in terms of form or scope. An example of each type of essay item follows.

Sample Extended-Response Essay Item

Explain the difference between the S-R (Stimulus-Response) and the S-O-R (Stimulus-Organism-Response) theories of personality. Include in your answer (a) brief descriptions of both theories, (b) supporters of both theories and (c) research methods used to study each of the two theories. (10 pts. 20 minutes)

Sample Short-Answer Essay Item

Identify research methods used to study the S-R (Stimulus-Response) and S-O-R (Stimulus-Response-Organism) theories of personality. (5 pts. 10 minutes).

Advantages in Using Essay Items

Essay items...

- ? are easier and less time consuming to construct than are most other item types.
- ? provide a means for testing student's ability to compose an answer and present it in a logical manner.
- ? can efficiently measure higher order cognitive objectives (e.g., analysis, synthesis, evaluation).

Limitations in Using Essay Items.

Essay items...

- ? cannot measure a large amount of content or objectives.
- ? generally provide low test and test scorer reliability.
- ? require an extensive amount of instructor's time to read and grade.
- ? generally do not provide an objective measure of student achievement or ability (subject to bias on the part of the grader).

SUGGESTIONS FOR WRITING ESSAY TEST ITEMS

1. Prepare essay items that elicit the type of behavior you want to measure.

Learning Objective: The student will be able to explain how the normal curve serves as a statistical model.

Undesirable: Describe a normal curve in terms of: symmetry, modality, kurtosis and skewness.

Desirable: Briefly explain how the normal curve serves as a statistical model for estimation and hypothesis testing.

2. Phrase each item so that the student's task is clearly indicated.

Undesirable: Discuss the economic factors which led to the stock market crash of 1929.

Desirable: Identify the three major economic conditions which led to the stock market crash of 1929. Discuss briefly each condition in correct chronological sequence and in one paragraph indicate how the three factors were interrelated.

3. Indicate for each item a point value or weight and an estimated time limit for answering.

Undesirable: Compare the writings of Bret Harte and Mark Twain in terms of settings, depth of characterization, and dialogue styles of their main characters.

*Desirable: Compare the writing of Bret Harte and Mark Twain in terms of settings, depth of characterization and dialogue styles of their main characters.
(10 points 20 minutes)*

4. Ask questions that will elicit responses on which experts could agree that one answer is better than another.
5. Avoid giving the student a choice among optional items as this greatly reduces the reliability of the test.
6. It is generally recommended for classroom examinations to administer several short-answer items rather than only one or two extended-response items.

SUGGESTIONS FOR SCORING ESSAY ITEMS

1. Choose a scoring model. Two of the more common scoring models are ANALYTICAL SCORING and GLOBAL QUALITY (Holistic).

ANALYTICAL SCORING: Each answer is compared to an ideal answer and points are assigned for the inclusion of necessary elements. Grades are based on the number of accumulated points either absolutely (i.e., A=10 or more points, B=6-9 pts., etc.) or relatively (A=top 15% of scores, B=next 30% of scores, etc.)

GLOBAL QUALITY: Each answer is read and assigned a score (e.g., grade, total points) based either on the total quality of the response or on the total quality of the response relative to other student answers.

Example Essay Item and Grading Models

“Americans are a mixed-up people with no sense of ethical values. Everyone knows that baseball is far less necessary than food and steel, yet they pay ball players a lot more than farmers and steelworkers.”

WHY? Use 3 - 4 sentences to indicate how an economist would explain the above situation

Analytical Scoring

<u>Necessary Elements to be Included in Response</u>	<u>Points</u>
Salaries are based on demand relative to supply of such services.	3
Excellent ball players are rare.	2
Ball clubs have a high demand for excellent players.	2
Clarity of Response	<u>2</u>
	9 pts.

Global Quality

Assign scores or grades on the overall quality of the written response as compared to an ideal answer. Or, compare the overall quality of a response to other student responses by sorting the papers into three stacks:

Below Average

Average

Above Average

Read and sort each stack again and divide into three more stacks.

Below Average

Average

Above Average

Below Average Above
Avg. Avg.

Below Average Above
Avg. Avg.

Below Average Above
Avg. Avg.

In total, nine discriminations can be used to assign test grades in this manner. The number of stacks or discriminations can vary to meet your needs.

2. Try not to allow factors which are irrelevant to the learning outcomes being measured affect your grading (i.e., handwriting, spelling, neatness).
3. Read and grade all class answers to one item before going on to the next item.
4. Read and grade the answers without looking at the students' names to avoid possible preferential treatment.
5. Occasionally shuffle papers during the reading of answers to help avoid any systematic order effects (i.e., Sally's "B" work always followed Jim's "A" work thus it looked more like "C" work). Use a numbering system to identify papers as opposed to student names so there is not bias from previous student work.
6. When possible, ask another instructor to read and grade your students' responses to see how closely the scores match.

PROBLEM SOLVING TEST ITEMS

Another form of a subjective test item is the problem solving or computational exam question. Such items present the student with a problem situation or task and require a demonstration of work procedures and a correct solution, or just a correct solution. This kind of test item is classified as a subjective type of item due to the procedures used to score item responses. Instructors can assign full or partial credit to either correct or incorrect solutions depending on the quality and kind of work procedures presented. An example of a problem solving test item follows.

Example Problem Solving Test Item

It was calculated that 75 men could complete a strip on a new highway in 70 days. When work was scheduled to commence, it was found necessary to send 25 men on another road project. How many days longer will it take to complete the strip? Show your work for full or partial credit.

Advantages in Using Problem Solving Items

Problem solving items...

- ? minimize guessing by requiring the students to provide an original response rather than to select from several alternatives.
- ? are easier to construct than are multiple-choice or matching items.
- ? can most appropriately measure learning objectives which focus on the ability to apply skills or knowledge in the solution of problems.
- ? can measure an extensive amount of content or objectives.

Limitations in Using Problem Solving Items

Problem solving items...

- ? generally provide low test and test scorer reliability.
- ? require an extensive amount of instructor time to read and grade.
- ? generally do not provide an objective measure of student achievement or ability (subject to bias on the part of the grader when partial credit is given).

SUGGESTIONS FOR WRITING PROBLEM SOLVING TEST ITEMS

1. Clearly identify and explain the problem.

Undesirable: *During a car crash, the car slows down at the rate of 490 m/sec^2 . What is the magnitude and direction of the force acting on a 100-kg driver?*

Desirable: *During a car crash, the car slows down at the rate of 490 m/sec^2 . Using the car as a frame of reference, what is the magnitude and direction of the gram force acting on a 100-kg driver?*

2. Provide directions which clearly inform the student of the type of response called for.

Undesirable: *An American tourist in Paris finds that he weighs 70 kilograms. When he left the United States he weighed 144 pounds. What was his net change in weight?*

Desirable: *An American tourist in Paris finds that he weighs 70 kilograms. When he left the United States he weighed 144 pounds. What was his net weight change in pounds?*

3. State in the directions whether or not the student must show his/her work procedures for full or partial credit.

Undesirable: *A double concave lens is made of glass with $n=1.50$. If the radii of curvature of the two lens surfaces are both 30.0 cm, what is the focal length of the lens?*

Desirable: *A double concave lens is made of glass with $n= 1.50$. If the radii of curvature of the two lens surfaces are both 30.0 cm, what is the focal length of the lens?
Show your work to receive full or partial credit.*

4. Clearly separate item parts and indicate their point values.

A man leaves his home and drives to a convention at an average rate of 50 miles per hour. Upon arrival, he finds a telegram advising him to return at once. He catches a plane that takes him back at an average rate of 300 miles per hour.

Undesirable: *If the total traveling time was $1 \frac{3}{4}$ hours, how long did it take him to fly back?
How far from his home was the convention?*

Desirable: *If the total traveling time was $1 \frac{3}{4}$ hours:
(1) How long did it take him to fly back? (1 pt.)
(2) How far from his home was the convention? (1 pt.)
Show your work for full or partial credit.*

5. Use figures, conditions and situations which create a realistic problem.

Undesirable: *An automobile weighing 2,840 N (about 640 pounds) is traveling at a speed of 300 miles per hour. What is the car's kinetic energy? Show your work. (2 pts.)*

Desirable: *An automobile weighing 14,200 N (about 3200 pounds) is traveling at a speed of 12m/sec. What is the car's kinetic energy? Show your work. (2 pts.)*

6. Ask questions that elicit responses on which experts could agree that one solution and one or more work procedures are better than others.

7. Work through each problem before classroom administration to double-check accuracy.

PERFORMANCE TEST ITEMS.

A performance test item is designed to assess the ability of a student to perform correctly in a *simulated situation* (i.e., a situation in which the student will be ultimately expected to apply his/her learning). The concept of *simulation* is central in performance testing; a performance test will simulate to some degree a real life situation to accomplish the assessment. In theory, a performance test could be constructed for any skill and real life situation. In practice, most performance tests have been developed for the assessment of vocational, managerial, administrative, leadership, communication, interpersonal and physical education skills in various simulated situations. An illustrative example of a performance test is provided below.

Sample Performance Test Item

Assume that some of the instructional objectives of an urban planning course include the development of the student's ability to effectively use the principles covered in the course in various "real life" situations common for an urban planning professional. A performance test item could measure this development by presenting the student with a specific situation which represents a "real life" situation. For example,

An urban planning board makes a last minute request for the professional to act as consultant and critique a written proposal which is to be considered in a board meeting that very evening. The professional arrives before the meeting and has one hour to analyze the written proposal and prepare his critique. The critique presentation is then made verbally during the board meeting; reactions of members of the board or the audience include requests for explanation of specific points or informed attacks on the positions taken by the professional.

The performance test designed to simulate this situation would require that the student to be tested role play the professional's part, while students or faculty act the other roles in the situation. Various aspects of the "professional's" performance would then be observed and rated by several judges with the necessary background. The ratings could then be used both to provide the student with a diagnosis of his/her strengths and weaknesses and to contribute to an overall summary evaluation of the student's abilities.

Advantages in Using Performance Test Items

Performance test items...

- ? can most appropriately measure learning objectives which focus on the ability of the students to apply skills or knowledge in real life situations.
- ? usually provide a degree of test validity not possible with standard paper and pencil test items.
- ? are useful for measuring learning objectives in the psychomotor domain.

Limitations in Using Performance Test Items.

Performance test items...

- ? are difficult and time consuming to construct.
- ? are primarily used for testing students individually and not for testing groups.
Consequently, they are relatively costly, time consuming, and inconvenient forms of testing.
- ? generally provide low test and test scorer reliability.
- ? generally do not provide an objective measure of student achievement or ability (subject to bias on the part of the observer/grader).

SUGGESTIONS FOR WRITING PERFORMANCE TEST ITEMS

1. Prepare items that elicit the type of behavior you want to measure.
2. Clearly identify and explain the simulated situation to the student.
3. Make the simulated situation as “life-like” as possible.
4. Provide directions which clearly inform the students of the type of response called for.
5. When appropriate, clearly state time and activity limitations in the directions.
6. Adequately train the observer(s)/scorer(s) to ensure that they are fair in scoring the appropriate behaviors.

III. ASSESSING TEST ITEM QUALITY

This section of the bulletin presents a self-review checklist of item quality. You can use the information from the checklist to identify strengths and weaknesses in your item writing.

CHECKLIST FOR EVALUATING TEST ITEMS

EVALUATE YOUR TEST ITEMS BY CHECKING THE SUGGESTIONS WHICH YOU FEEL YOU HAVE FOLLOWED.

Multiple-Choice Test Items

- When possible, stated the stem as a direct question rather than as an incomplete statement.
- Presented a definite, explicit and singular question or problem in the stem.
- Eliminated excessive verbiage or irrelevant information from the stem.
- Included in the stem any word(s) that might have otherwise been repeated in each alternative.
- Used negatively stated stems sparingly. When used, underlined and/or capitalized the negative word(s).
- Made all the alternatives plausible and attractive to the less knowledgeable or skillful student.
- Made the alternatives grammatically parallel with each other, and consistent with the stem.
- Made the alternatives mutually exclusive.
- When possible, presented alternatives in some logical order (e.g., chronologically, most to least).
- Made sure there was only one correct or best response per item.
- Made alternatives approximately equal in length.
- Avoided irrelevant clues such as grammatical structure, well known verbal associations or connections between stem and answer.
- Used at least four alternatives for each item.
- Randomly distributed the correct response among the alternative positions throughout the test having approximately the same proportion of alternatives a, b, c, d, and e as the correct response.
- Used the alternatives “none of the above” and “all of the above” sparingly. When used, such alternatives were occasionally the correct response.

True-False Test Items

- _____ Based true-false items upon statements that are absolutely true or false, without qualifications or exceptions.
- _____ Expressed the item statement as simply and as clearly as possible.
- _____ Expressed a single idea in each test item.
- _____ Included enough background information and qualifications so that the ability to respond correctly did not depend on some special, uncommon knowledge.
- _____ Avoided lifting statements from the text, lecture or other materials.
- _____ Avoided using negatively stated item statements.
- _____ Avoided the use of unfamiliar language.
- _____ Avoided the use of specific determiners such as “all,” “always,” “none,” “never,” etc., and qualifying determiners such as “usually,” “sometimes,” “often,” etc.
- _____ Use more false items than true items (but not more than 15% additional false items).

Matching Test Items

- _____ Included directions which clearly stated the basis for matching the stimuli with the response.
- _____ Explained whether or not a response could be used more than once and indicated where to write the answer.
- _____ Used only homogeneous material.
- _____ When possible, arranged the list of responses in some systematic order(e.g., chronologically, alphabetically).
- _____ Avoided grammatical or other clues to the correct response.
- _____ Kept items brief (limited the stimuli to under 10).
- _____ Included more responses than stimuli.
- _____ When possible, reduced the amount of reading time by including only short phrases or single words in the response list.

Completion Test Items

- _____ Omitted only significant words from the statement.
- _____ Did not omit so many words from the statement that the intended meaning was lost.
- _____ Avoided grammatical or other clues to the correct response.
- _____ Included only one correct response per item.

- _____ Made the blanks of equal length.
- _____ When possible, deleted the words at the end of the statement after the student was presented with a clearly defined problem.
- _____ Avoided lifting statements directly from the text, lecture or other sources.
- _____ Limited the required response to a single word or phrase.

Essay Test Items

- _____ Prepared items that elicited the type of behavior you wanted to measure.
- _____ Phrased each item so that the student's task was clearly indicated.
- _____ Indicated for each item a point value or weight and an estimated time limit for answering.
- _____ Asked questions that elicited responses on which experts could agree that one answer is better than others.
- _____ Avoided giving the student a choice among optional items.
- _____ Administered several short-answer items rather than 1 or 2 extended-response items.

Grading Essay Test Items

- _____ Selected an appropriate grading model.
- _____ Tried not to allow factors which were irrelevant to the learning outcomes being measured to affect your grading (e.g., handwriting, spelling, neatness).
- _____ Read and graded all class answers to one item before going on to the next item.
- _____ Read and graded the answers without looking at the student's name to avoid possible preferential treatment.
- _____ Occasionally shuffled papers during the reading of answers.
- _____ When possible, asked another instructor to read and grade your students' responses.

Problem Solving Test Items

- _____ Clearly identified and explained the problem to the student.
- _____ Provided directions which clearly informed the student of the type of response called for.
- _____ Stated in the directions whether or not the student must show work procedures for full or partial credit.
- _____ Clearly separated item parts and indicated their point values.
- _____ Used figures, conditions and situations which created a realistic problem.

- _____ Asked questions that elicited responses on which experts could agree that one solution and one or more work procedures are better than others.
- _____ Worked through each problem before classroom administration.

Performance Test Items

- _____ Prepared items that elicit the type of behavior you wanted to measure.
- _____ Clearly identified and explained the simulated situation to the student.
- _____ Made the simulated situation as “life-like” as possible.
- _____ Provided directions which clearly inform the students of the type of response called for.
- _____ When appropriate, clearly stated time and activity limitations in the directions.
- _____ Adequately trained the observer(s)/scorer(s) to ensure that they were fair in scoring the appropriate behaviors.

IV. REFERENCES FOR FURTHER READING

Ahmann, J.S. & Glock, M.D. (1981). Evaluating student progress. Boston, MA: Allyn and Bacon.

Gronlund, N.E. (1988). How to construct achievement tests. Englewood Cliffs, New Jersey: Prentice-Hall, Inc.

Gronlund, N.E. (1976). Measurement and evaluation in teaching (4th ed.), Chapters 5 - 10. New York: Macmillan Publishing Co.

Hopkins, C.D. and Antes, R.L., (1979). Classroom Testing: Construction. Itasca, IL: F.E. Peacock Publishers

Lindquist, E.F. (Ed.) (1951). Educational measurement. Washington, D.C.: American Council on Education. Chapter 7, Writing the Test Item.

Linn, R.L., (1989). Educational measurement (3rd ed.). New York: American Council on Education/Macmillan Publishing Co.

Mehrens, W.A. and Lehmann (1984). Measurement and evaluation in education and psychology. (3rd ed.). New York: Holt, Rinehart and Winston. Chapters 4 - 7.

Thorndike, R.L. (Ed.) (1971). Educational measurement (2nd ed.). Washington, D.C.: American Council on Education, Chapter 9 (Performance testing) and Chapter 10 (Essay exams).