

**APPENDIX**

**Using, Writing, and Evaluating  
Forced-Choice Assessments**

Excerpted and Adapted From

**IMPROVING YOUR TEST QUESTIONS**

John C. Ory  
Office of Instructional  
Resources  
University of Illinois  
at Urbana-Champaign

1985

# Suggestions for Using and Writing Forced-Choice Assessments

## 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*
- \*correct response*

### 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.
- measurement of student achievement or ability which does not require rater judgment.
- a wide sampling of content or learning outcomes.
- 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:** *The science of mind and behavior is 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?*





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. If you have decided to use a traditional single correct answer format, be sure there is only one correct or best answer.

**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. 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. A 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*  
(connection between stem and answer clue)

- a. length of the reservoir behind the dam.
- b. the volume of water behind the dam.
- \*c. the height of water behind the dam.
- d. the 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.

### True-False Test Items

A true-false item can be written in one of three forms: simple, complex, or compound. Answers can consist of only two choices (simple), more than two choices (complex), or two choices plus a conditional completion response (compound). An example of each type of true-false item follows:

**Sample True-False Item: Simple**

*The acquisition of character is a developmental process.*                      *True*                      *False*

**Sample True-False Item: Complex**

*The acquisition of character is a developmental process.*                      *True*                      *False*                      *Opinion*



### Sample True-False Item: Compound

*The acquisition of character is a developmental process.*                      *True*                      *False*

*If this statement is false, what makes it false?*

---

---

### Advantages in Using True-False Items

True-false items can provide

- the widest sampling of content or objectives per unit of testing time.
- scoring efficiency and accuracy.
- versatility in measuring all levels of cognitive ability.
- highly reliable test scores.
- an objective measurement of student achievement or ability.

### Limitations in Using True-False Items

True-false items

- incorporate an extremely high guessing factor. For simple true-false items, each student has a 50/50 chance of correctly answering the item without any knowledge of the item's content.
- can often lead an instructor to write ambiguous statements due to the difficulty of writing statements which are unequivocally true or false.
- do not discriminate between students of varying ability as well as other item types.
- can often include more irrelevant clues than do other item types.
- can often lead an instructor to favor testing of trivial knowledge.

### 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 built and 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 not 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 justice.*

7. Avoid the use of unfamiliar vocabulary.

**Undesirable:** *According to some politicians, the *raison d 'etre* 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 student 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°. (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.*

##### Factual Statements

  1. Fossils of primates first appear in the Ceno-  
zoic rock strata, while tribolite remains  
are found in the Protozoic rocks.

  2. The Arctic and Antarctic regions are  
sparsely populated.

  3. Plants have no nervous system.

  4. Large coal beds exist in Alaska.

##### Principles

a. There have been profound changes in the  
climate on earth.

b. Coordination and integration of action is generally  
slower in plants than in animals.

c. There is an increasing complexity of structure  
and functions from lower to higher forms of  
life.

d. All life comes from life and produces its own  
kind of living organism.

e. Light is a limiting factor to life.



**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
- 2. \_\_\_ Salt
- 3. \_\_\_ Ammonia
- 4. \_\_\_ Sulfuric Acid

- A.  $H_2SO_4$
- B.  $HCl$
- C.  $NaCl$
- D.  $H_2O$
- E.  $H_2HCl$

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.

**Column I**

**Undesirable**

**Desirable**

**Column II**

- \_\_\_ 1. Hunting for reasons to support one's beliefs.
- \_\_\_ 2. Accepting the values and norms of others as one's own even when they are contrary to previously held values.
- \_\_\_ 3. Attributing to others one's own unacceptable impulses, thoughts and desires.
- \_\_\_ 4. Ignoring disagreeable situations, topics, sights.

- |   |   |
|---|---|
| <ul style="list-style-type: none"> <li>a. Rationalization</li> <li>b. Identification</li> <li>c. Projection</li> <li>d. Introjection</li> <li>e. Denial of reality</li> </ul> | <ul style="list-style-type: none"> <li>a. Denial of reality</li> <li>b. Identification</li> <li>c. Introjection</li> <li>d. Projection</li> <li>e. Rationalization</li> </ul> |
|---|---|

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

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

- |  |  |
|--|--|
| <ul style="list-style-type: none"> <li>___ 1. Igneous rocks are reformed</li> <li>___ 2. The formation of coal requires</li> <li>___ 3. A geode is filled</li> <li>___ 4. Feldspar is classified as</li> </ul> | <ul style="list-style-type: none"> <li>A. a hardness of 7.</li> <li>B. with crystalline rock.</li> <li>C. a metamorphic rock.</li> <li>D. heat and pressure.</li> <li>E. through the solidification of molten lava.</li> </ul> |
|--|--|

**Desirable:** Avoid sentence completion due to grammatical clues.

- □
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.
- can efficiently measure lower levels of cognitive ability.
- can minimize guessing as compared to multiple-choice or true-false items.
- can usually provide a reliable 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 prepared.







- \_\_\_ 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.
- \_\_\_ Used 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 list of 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.