

# **A GUIDE TO PREPARING MULTIPLE-CHOICE ITEMS**

Created by Gerald Rosen, Ph.D.

## **INTRODUCTION**

The most commonly used type of test item in admissions, certification, and licensure examinations is the multiple-choice item. The purpose of this brochure is to assist test developers and representatives of professional associations in becoming more skilled in writing multiple choice test items. Preparing multiple-choice items may appear on the surface to be a relatively simple task, given familiarity with the subject matter. However, experienced test authors find that this task actually requires a great deal of skill, patience, and creativity.

In the instructional sections that follow, general knowledge is used as the content of the examples, in order to facilitate understanding of the general principles for professionals from diverse content areas. The skills needed for writing quality items are similar regardless of content discipline.

Like any other skill, multiple-choice item writing requires a significant amount of painstaking practice with appropriate feedback. It is essential that you try out your multiple-choice items on others. This process will help you to uncover ambiguities in wording and unintended violations of item-writing principles. Feedback from others will greatly enhance your item writing quality and productivity.

## **PARTS OF THE MULTIPLE-CHOICE ITEM**

All multiple-choice items consist of two basic parts: the **stem** and the **responses**:

The **stem** is the introductory statement or question that elicits the correct answer. The **responses** are suggested answers which complete the statement or answer the question asked in the stem, only one of which is the correct answer. In the examples that follow, four responses (i.e., 1 correct response and 3 incorrect responses) are presented. *Refer to the blank item form included with this booklet for guidance regarding the appropriate number of responses for the items that you prepare.*

**STEM** Who is the scientist most closely associated with the discovery of polio vaccine?

- RESPONSES**
1. Jonas Salk
  2. Louis Pasteur
  3. Edward Jenner
  4. Robert Koch

In the example above, of the four responses, the three incorrect responses are called **distractors** and the correct response is called the **answer**. One of the key elements in multiple-choice item writing is to be able to write high-quality distractors that effectively discriminate those who have mastered the material being assessed from those who have not.

Item stems can be written in the form of an incomplete statement as well as in question form. Thus, the item above could be modified as follows:

- The polio vaccine was discovered by:
1. Jonas Salk.
  2. Louis Pasteur.
  3. Edward Jenner.
  4. Robert Koch.

Note that the responses for this item are followed by a period, as the response serves to complete a sentence.

It is wise to draft an item first as a direct question, then revise it to an incomplete sentence if this offers a smoother, less repetitive wording.

### **CORRECT- VS. BEST-ANSWER ITEMS**

It is important for test item writers to keep in mind the distinction between items in which **only one** response is correct, and items in which more than one response may be technically correct but in which only one response is clearly the best answer.

#### **Correct-Answer Item**

The capital of Kentucky is:

1. Frankfort.
2. Lexington.
3. Louisville.
4. Ashland.

#### **Best-Answer Item**

The *major* goal of raising the legal drinking age is to:

1. reduce driving fatalities and injuries.
2. reduce liquor consumption.
3. discourage the habit of drinking by adolescents.
4. reduce roadside litter.

In writing best-answer items, care must be taken to word the stem in such a way so as to make it clear that only one response is preferred. In the above example, the word **major** achieves this objective.

## REFINING MULTIPLE-CHOICE ITEMS

The following principles are essential in developing high-quality multiple-choice items that neither confuse nor penalize the examinee nor reward the examinee by inadvertently providing clues which make the correct answer obvious. As you acquire experience writing items, you should refer to these principles frequently. Apply these principles when evaluating the items written by others.

1. When writing test items, always **put the correct answer as the first response** so that reviewers can identify the correct answer easily.
2. The item stem should be clear enough to provide the examinee with sufficient information to anticipate the type of answer before looking at the responses.

### Poor

Diabetes is:

1. a disorder associated with the body's absorption of glucose.
2. a digestive disorder which afflicts adults.
3. a potentially fatal disorder if not detected early.
4. a disease that can result in blindness.

### Better

Diabetes is a chronic disease characterized by:

1. the inability of the body to absorb glucose.
2. a genetic abnormality resulting in high blood sugar levels.
3. an excessive ingestion of sweet foods during infancy.
4. an inability to digest amino acids.

Note that in the above example; making the stem more specific clarifies the example significantly.

3. Whenever possible, include in the stem all words that would otherwise have to be repeated in each of the responses.

### Poor

The process whereby amniotic fluid is removed removed from a pregnant woman to test for possible defects:

1. is known as amniocentesis.
2. is known as cesarean section.
3. is known as embryonic analysis.
4. is known as fetal catheterization.

### Better

The process whereby amniotic fluid is from a pregnant woman to test for possible birth defects is known as:

1. amniocentesis.
2. cesarean section.
3. embryonic analysis.
4. fetal catheterization.

4. It is best to avoid negatively stated item stems - i.e., items that require the examinee to select an incorrect response rather than a correct response. These types of stems tend to confuse the examinee and thus tend to make the items less effective in assessing subject matter knowledge and understanding.

### Poor

### Better

Which of the following authors is **not** a major American novelist?

1. George Orwell
2. Saul Bellow
3. John Steinbeck
4. Ernest Hemingway

Which of the following authors is a major British novelist?

1. George Orwell
2. Saul Bellow
3. John Steinbeck
4. Ernest Hemingway

5. In a multiple-choice item, the correct response must clearly stand out as the one that experts in the field would recognize as the best response. When there is doubt or controversy about the correct response, not only is the examinee confused, but the item could be challenged as being inappropriate.

**Poor**

The primary cause of the Vietnam War was:

1. Vietnamese nationalism.
2. French imperialism.
3. Chinese expansionism.
4. American imperialism.

**Better**

Which of the following factors most likely sustained the military campaign of the Vietnamese insurgents in the Vietnam War?

1. Nationalism
2. Chinese expansionism
3. Soviet militarism
4. American expansionism

6. All responses should be grammatically consistent with the item stem, and all response options should be parallel. Lack of parallelism can sometimes make it possible for examinees to choose the intended answer even when they do not have the skill or knowledge that an item is intended to measure.

**Poor**

An American city noted for its production of automobiles is:

1. Detroit.
2. Ohio.
3. Indiana.
4. Philadelphia.

**Better**

An American city noted for its production of automobiles is:

1. Detroit.
2. Cincinnati.
3. Indianapolis.
4. Philadelphia.

**Poor**

Serious communicable diseases that can cause developmental disability to the offspring of a pregnant mother are:

1. rubella and chickenpox.

**Better**

A serious communicable disease that can cause developmental disability to the offspring of a pregnant mother is:

1. rubella.

2. cancer.
3. rheumatism and arthritis.
4. sinus infection.

2. cancer.
3. rheumatism.
4. sinus infection.

Note how the use of a **singular** stem necessitates the use of a singular response.

7. Distractors should represent unsafe practices or commonly held misconceptions and should be plausible. Avoid using distractors which even the most uninformed examinee would recognize as being incorrect. **Also, the use of humorous or absurd distractors is not appropriate in standardized test items.**

**Poor**

**Better**

The person who developed the psychoanalytic theory of human behavior is:

The person who developed the psychoanalytic theory of human behavior is:

1. Sigmund Freud.
2. Babe Ruth.
3. Jimmy Carter.
4. Burt Reynolds.

1. Sigmund Freud.
2. B. F. Skinner.
3. Carl Rogers.
4. John Watson.

8. Avoid the use of language which is unnecessarily technical or unfamiliar to appropriately trained examinees. Always take into account the relative reading level and level of knowledge of people taking the test.

**Poor**

**Better**

A biological explanation for sex difference in human behavior would most likely be supported by research on:

A biological explanation for sex difference in human behavior would most likely be supported by research on:

1. fetally androgenized females.
2. cultural conditioning.
3. pseudohermaphrodites.
4. neuroendocrinology.

1. sex-hormone-level disorders during the fetal stage.
2. cultural stereotypes of males and females.
3. people who have both male and female physical characteristics.
4. comparisons of male and female athletic performance.

9. Avoid phrasing the correct answer to a question directly from a textbook. Such answers are usually more technical than the distractors and can be easily answered by the examinee, with little or no content knowledge.

**Poor**

**Better**

What is anorexia?

1. A psychophysiological condition of post-pubescent adolescent whereby caloric intake is reduced
2. A prolonged period of nausea
3. An adolescent growth disorder
4. A disorder of the pituitary gland

What is anorexia?

1. Lack or loss of the appetite for foods
2. A prolonged period of nausea
3. An adolescent growth disorder
4. A disorder of the pituitary gland

10. Responses should not overlap or include each other so that the examinee could possibly select more than one response and still be technically correct.

**Poor**

The right to vote in the United States is granted to individuals of what age?

1. 18
2. 16
3. 17
4. 19

**Better**

The right to vote in the United States is granted to individuals **beginning** at what age?

1. 18
2. 16
3. 17
4. 19

Note that in the example on the left, both number 1 and number 4 could be considered correct.

11. Do not include the response "**all of the above**" or "**none of the above**". The flaw in these items is that it is logically difficult to claim that one of the other responses in the item is either correct or incorrect. When "**none of the above**" is the correct answer, it still cannot be determined whether the examinee knows what the correct answer is only that he or she recognized wrong answers.
12. If the item calls for an evaluation or judgment, be sure to provide the authority or source of the judgment. This is especially important when dealing with ethical decisions and decisions regarding interpersonal relationships or controversial subject matter.

**Poor**

The best way to discipline a child is to:

1. reward good behavior.
2. use physical punishment for serious offenses.
3. explain to the child why an undesirable behavior displeases you.
4. ignore misbehavior.

**Better**

According to behavioral psychology, the best way to discipline a child is to:

1. reward good behavior.
2. use physical punishment for serious offenses.
3. explain to the child why an undesirable behavior displeases you.
4. ignore misbehavior.

13. Do not use the form “**what would you do**” or “**what do you believe?**” In items of this kind, the examinee could always defend his or her response selection as being correct, since each person has a unique frame of reference.
14. Item answers should not be worded so that they might be out of date at the time of testing.

**Poor**

In the last presidential election, a vice-presidential candidate who was defeated was:

1. Walter Mondale.
2. Robert Dole.
3. Sargent Shriver.
4. George Bush.

**Better**

In the presidential election of 1980, a vice-presidential candidate who was defeated was:

1. Walter Mondale.
2. Robert Dole.
3. Sargent Shriver.
4. George Bush.

15. Items should be written to assess knowledge of meaningful facts and concepts, not trivial information.

**Poor**

The assassination of President John F. Kennedy occurred on what date?

1. November 22, 1963
2. December 4, 1961
3. February 18, 1961
4. June 7, 1959

**Better**

President John F. Kennedy was assassinated:

1. before the passage of the major civil rights legislation of the 1960's.
2. before the Bay of Pigs invasion of Cuba.
3. after the Vietnam war ended.
4. before the construction of the Berlin Wall.

Note how the item on the right assesses the sequence of historical events rather than a single date.

16. Do not include any inessential information in the item stem. Stems should ask the question as succinctly as possible.

**Poor**

In 1850, Adolphe Chattin, then professor of pharmacy in Paris, believed that goiter resulted from an inadequate amount of iodine in the diet. The thyroid is part of which body system?

1. Endocrine
2. Nervous
3. Locomotor
4. Cardiovascular

**Better**

The thyroid is part of which body system?

1. Endocrine
2. Nervous
3. Locomotor
4. Cardiovascular

17. Avoid providing the examinee with clues that will make it easy for him or her to detect the correct answer.
  - a. Do not make the correct answer much shorter, longer, or more technical than the distractors. If this is not possible, try to have two relatively short and two relatively long distractors or two relatively technical and two relatively non-technical distractors.
  - b. Do not put the same key words or descriptive words both in the stem and in the correct answer but not in the distractors.
  - c. Do not make the correct answer clear and concise and the distractors vague and ambiguous.
18. The item stem should be specific enough to pose only one question or problem, and each distractor should be related to that question or problem. It is very confusing for the examinee to read an ambiguous stem for which several distractors could be construed as being correct.

**Poor**

Alcoholic beverage consumption is believed to be:

1. a tradition in Western cultures.
2. a vice by those with certain religious beliefs.
3. a means to achieve desirable states of relaxation.
4. a major cause of family conflict and disorganization.

**Better**

Data on per capita alcoholic beverage consumption in the United States during the last 10 years suggest what statistical trend?

1. Increased consumption by females
2. Decreased consumption by adolescents
3. Abstinence among all age groups
4. Abstinence among those over age 65

19. For certain tests, items which assess an examinee's response to a job-related situation are required. These items are difficult to develop because it is often hard to write distractors that result in an item that is neither too easy nor too difficult. One suggestion for writing good job-related situation items is to first write a question asking the best way a professional would respond to a given situation. Then prepare a list of responses with a wide range of acceptability among experienced professionals in the appropriate occupational group. Finally, present the question and list of responses to several professionals from the occupational group and have them select what they believe is the best answer, and three reasonable distractors. While this is a time-consuming practice, it is essential for writing items of this kind. It is also important that the stem in these items provide only essential information. Superfluous detail should be avoided.

20. In certain situations, it is useful to base a group of items on one information set. An information

set can present a context or relate questions to commonly encountered situations. If this is done, make sure that the information presented in one item does not result in giving away the correct answer for a related item.

### **INFORMATION SET**

QUESTIONS 1 THROUGH 3 REFER  
TO THE FOLLOWING INFORMATION:

On a biology examination, the following scores were obtained by students:

90  
88  
85  
76  
72

1. What is the average or mean score for this group of students?

1. 82.2
2. 85.0
3. 79.5
4. 84.8

2. What is the median score for this group of students?

1. 85
2. 88
3. 76
4. 90

3. What is the range of the scores for this group of students?

1. 18 points
2. 90 points
3. 72 points
4. 4 points

21. When possible, avoid the use of gender pronouns such as he, she, his, her, etc. By rewording the

stem and/or responses to eliminate gender, neutral items can be constructed.

**Poor**

If an elementary school teacher has a child who has been stung by a bee, she should do which of the following things first?

1. Determine if he is allergic to bee stings.
2. Remove the stinger from her with tweezers.
3. Attempt to kill the bee.
4. Bandage the affected.

**Better**

Of the following, what is the first thing an elementary school teacher should do to treat a child who has been stung by a bee?

1. Determine if the child is allergic to bee stings.
2. Remove the stinger with tweezers.
3. Attempt to kill the bee.
4. Bandage the affected area.

22. An effort should be made to assess higher-level cognitive skills such as application of knowledge and problem solving. While it is sometimes difficult to write items of this type, they are needed to assess important components of examinee competence.

**Knowledge**

Compared to boys, how much earlier or later is the average start of puberty for girls?

1. 2 years earlier for girls
2. 3 years later for girls
3. 1 year earlier for girls
4. 2 years later for girls

**Application of Knowledge**

A chaperone at a seventh-grade dance would generally observe which of the following conditions?

1. Girls are taller than boys.
2. Boys are taller than girls.
3. Girls and boys are of approximately equal height.
4. Most of the boys have completed puberty.

**Problem Solving**

If a school principal wanted to arrange for a dance where the boys and girls would be of approximately equal height, this could be best achieved by inviting:

1. seventh-grade girls and eighth-grade boys.
2. seventh-grade boys and ninth-grade girls.
3. ninth-grade boys and ninth-grade girls.
4. eighth-grade girls and seventh-grade boys.

23. Items must be free of language, descriptions or terminology that could reinforce common stereotypes concerning any segment of the candidate population. Avoid casting individuals in stereotyped roles. For example, business executives, doctors, lawyers, homemakers and care-givers can all be male or female.

24. Avoid the use of language or descriptions that might be offensive to any segment of the candidate population such as members of a particular religion, ethnic group, social class, racial minority, etc.
25. Avoid the use of terms, descriptions, language or stereotyped roles that might be offensive to individuals with disabilities or cast them in a negative light by virtue of the disability.
26. Ensure that the context, setting and content of items are equally appropriate for and familiar to all segments of the candidate population including minorities.

### Checklist for Submission of Multiple-Choice Items

1.	Do all items present the appropriate number of responses per item? (See item form for this information.)	YES	NO
2.	Is each item typed (double-spaced) or written clearly on a separate item form?	YES	NO
3.	Is the correct answer the first option and is it marked with an asterisk?	YES	NO
4.	In the lower right-hand corner of the item form, is the area of the content outline specified?	YES	NO
5.	Is each item worded clearly and concisely and is each response grammatically parallel with the item stem?	YES	NO
6.	Have items with " <b>all of the above</b> " or " <b>none of the above</b> " responses been eliminated?	YES	NO
7.	Does each item stem address only one problem or content area?	YES	NO
8.	Where the correct answer requires units of measurement, has the unit name been provided?	YES	NO
9.	Are all item distractors plausible to the examinee?	YES	NO
10.	Has all excessively technical language or jargon been eliminated from items?	YES	NO
11.	Where the correct answer is based on a specific authority or source, has this been cited?	YES	NO
12.	Has the use of negatively stated item stems been avoided as much as possible?	YES	NO
13.	Are there items that assess application and problem solving, as well as items that assess knowledge?	YES	NO
14.	Has language that might stereotype or be offensive to minorities, women, persons with disabilities, etc. been removed?	YES	NO

**References for additional material on item development:**

Brown, F.G. (1983) **Principles of Educational and Psychological Testing**. Third edition. New York: Holt, Rinehart.

Lindeman, R.H., & Merenda, R.F. (1979) **Educational Measurement**. Second edition. Glenville, IL: Scott, Foresman.

Wesman, A.G. (1971) Writing the Test Item. In R.L. Thomdike (Ed.), **Educational Measurement**. Washington: American Council on Education, Pages 81 - 119.