

EIGHTH EDITION

***EDUCATIONAL TESTING
AND MEASUREMENT:***
Classroom Application and Practice

2007

TOM KUBISZYN

University of Houston

GARY BORICH

The University of Texas at Austin



JOHN WILEY & SONS, INC.

MULTIPLE-CHOICE ITEMS

Another popular item format is the multiple-choice question. Practically everyone has taken multiple-choice tests at one time or another, but probably more often in high school and college than in elementary school. This doesn't mean that multiple-choice items are not appropriate in the elementary years; it suggests only that one needs to be cautious about using them with younger children.

Multiple-choice items are unique among objective test items because, contrary to popular opinion, they enable you to measure at the higher levels of the taxonomy of educational objectives. Our discussion of multiple-choice items will be in two parts. The first part will consider the mechanics of multiple-choice item construction applied to knowledge level questions. The second part will deal with the construction of higher level multiple-choice items. As before, let's start by using common sense to identify good and poor multiple-choice items in the following exercise.

Exercise: Place a *G* in the space next to a good item and a *P* next to a poor item.

- _____ 1. U.S. Grant was an
 - a. president.
 - b. man.
 - c. alcoholic.
 - d. general.
- _____ 2. In what year did humans first set foot on the moon?
 - a. 1975
 - b. 1957
 - c. 1969
 - d. 1963
- _____ 3. The free-floating structures within the cell that synthesize protein are called
 - a. chromosomes.
 - b. lysosomes.
 - c. mitochondria.
 - d. free ribosomes.

- _____4. The principal value of a balanced diet is that it
- increases your intelligence.
 - gives you something to talk about with friends.
 - promotes mental health.
 - promotes physical health.
 - improves self-discipline.
- _____5. Some test items
- are too difficult.
 - are objective.
 - are poorly constructed.
 - have multiple defensible answers.
- _____6. Which of the following are not associated with pneumonia?
- quiet breathing
 - fever
 - clear chest x-ray
 - a and c
 - b and c
- _____7. When 53 Americans were held hostage in Iran,
- the United States did nothing to try to free them.
 - the United States declared war on Iran.
 - the United States first attempted to free them by diplomatic means and later attempted a rescue.
 - the United States expelled all Iranian students.
- _____8. The square root of 256 is
- 14.
 - 16.
 - 4×4
 - both a and c.
 - both b and c.
 - all of the above.
 - none of the above.
- _____9. When a test item and the objective it is intended to measure match in learning outcome and conditions, the item
- is called an objective item.
 - has content validity.
 - is too easy.
 - should be discarded.

Go over the exercise again. Chances are you'll find a few more problems the second time. Here's the answer key, and a breakdown of the faults found in each item follows in the text.

Answers: 1. P; 2. G; 3. P; 4. P; 5. P; 6. P; 7. P; 8. P; 9. G.

Most students would probably pick up on the grammatical clue in the first item. The article "an" eliminates options a, b, and d immediately, since "U.S. Grant was an man," "an president," or "an general" are not grammatically correct statements. Thus option c is the only option that forms a grammatically correct sentence. Inadvertently providing students with grammatical clues to the correct answer is very common in multiple-choice items. The result is decreased test validity. Students can answer items correctly because of knowledge of grammar, not content.

Replacing "an" with "a/an" would be one way to eliminate grammatical clues in your own writing. Other examples would be "is/are," "was/were," "his/her," and so on. As an alternative, the article, verb, or pronoun may be included in the list of options, as the following example illustrates:

Poor: Christopher Columbus came to America in a

- a. car.
- b. boat.
- c. airplane.
- d. balloon.

Better: Christopher Columbus came to America in

- a. a car.
- b. a boat.
- c. an airplane.
- d. a balloon.

Let's return to the first item and replace "an" with "a/an":

U.S. Grant was a/an

- a. president.
- b. man.
- c. alcoholic.
- d. general.

There! We've removed the grammatical clue, and we now have an acceptable item, right? Not quite. We now have an item free of grammatical clues, but it is still seriously deficient. What is the correct answer?

This item still has a serious flaw: multiple defensible answers. In fact, all four options are defensible answers! U.S. Grant was a president, a man, a general, and, as historians tell us, an alcoholic. Including such an item on a test would contribute nothing to your understanding of student knowledge. But what can you do when you have an item with more than one defensible answer? The answer, of course, is to eliminate the incorrect but defensible option or options.

Let's assume item 1 was written to measure the following objective:

The student will discriminate among the U.S. presidents immediately before, during, and immediately after the U.S. Civil War.

We could modify item 1 to look like this:

U.S. Grant was a

- a. general.
- b. slave.
- c. pirate.
- d. trader.

This item is fine, from a technical standpoint. The grammatical clue has been eliminated and there is but one defensible answer. However, it does not match the instructional objective; it is not very valuable as a measure of student achievement of the objective.

We could also modify the item to look like this:

Of the following, who was elected president after the Civil War?

- a. U.S. Grant
- b. Andrew Johnson
- c. Abraham Lincoln
- d. Andrew Jackson

This item is technically sound, and all response alternatives are relevant to the instructional objective. It meets the two main criteria for inclusion in a test: The item is technically well constructed and it matches the instructional objectives.

We said item 2 was good, but it can still stand some improvement. Remember when we recommended arranging lists for matching items in alphabetical or chronological order? The same holds true for multiple-choice items. To make a good item even better, arrange the options in chronological order. Revised, the item should look like this:

In what year did humans first set foot on the moon?

- a. 1957
- b. 1963
- c. 1969
- d. 1975

The major deficiency in item 3 is referred to as a "stem clue." The statement portion of a multiple-choice item is called the stem, and the correct answer and incorrect choices are called options or response alternatives. A stem clue occurs when the same word or a close derivative occurs in both the stem and options, thereby clueing the test taker to the correct answer. In item 3 the word free in the option is identical to free in the stem. Thus the wise test taker has a good chance of answering the item correctly without mastery of the content being measured. This fault can be eliminated by simply rewording the item without the word free.

The structures within the cell that synthesize protein are called

- a. chromosomes.
- b. lysosomes.
- c. mitochondria.
- d. ribosomes.

Item 4 is related to the "opinionated" items we considered when we discussed true-false items. Depending on the source, or referent, different answers may be the "right" answer. To Person X, the principal value may be to promote physical health; to Person Y, the principal value may be to improve self-discipline. As stated earlier, when you are measuring a viewpoint or opinion, be sure to state the referent or source. To be acceptable the item should be rewritten to include the name of an authority:

The USDA says the principal value of a balanced diet is that it

- a. increases your intelligence.
- b. gives you something to talk about.
- c. promotes mental health.
- d. promotes physical health.
- e. improves self-discipline.

Item 5 is, of course, meaningless. It has at least two serious faults. To begin with, the stem fails to present a problem, and it fails to focus the item. What is the item getting at? The test taker has no idea what to look for in trying to discriminate among the options. The only way to approach such an item is to look at each option as an individual true-false item. This is very time-consuming and frustrating for the test taker. Be sure to focus your multiple-choice items by presenting a problem or situation in the stem.

Like item 1, item 5 also has more than one defensible answer. However, option d seems to control this problem. But if more than a single option is defensible, how can you mark as incorrect someone who chooses a, b, or c and not d? Sometimes, however, you may wish to construct items that have two defensible answers. Is there any way to avoid the problem just mentioned? Fortunately, there is a way to avoid the problem, as illustrated in item 6:

Which of the following are not associated with pneumonia?

- a. quiet breathing
- b. fever
- c. clear chest x-ray
- d. a and c
- e. b and c

Where the possibility of more than one answer is desirable, use an option format like that just shown. This approach avoids the wording problems we ran into in item 5. We would caution, however, that "a and b," "b and c," and so on should be used sparingly.

Now, how about the rest of item 6; is it okay? No; again a grammatical clue is present. The word *are* indicates a plural response is appropriate. Options a, b, and c can automatically be eliminated, leaving the test taker with a 50% chance of guessing correctly. This fault can be corrected by using the same approach we used with item 1, where we substituted "a/an" for "an." Of course, in this instance we would substitute "is/are" for "are." Rewritten, the item looks like this:

Which of the following is/are not associated with pneumonia?

- a. quiet breathing
- b. fever
- c. clear chest x-ray
- d. a and c
- e. b and c

All set? Not yet! Remember what we said about negatives? Let's highlight the "not" with uppercase letters, italics, or underlining to minimize the likelihood of someone misreading the item. After this revision we have an acceptable multiple-choice item.

Two very common faults in multiple-choice construction are illustrated by item 7. First, the phrase "the United States" is included in each option. To save space and time, add it to the stem. Second, the length of options could be a giveaway. Multiple-choice item writers have a tendency to include more information in the correct option than in the incorrect options. Test-wise students take advantage of this tendency, since past

experience tells them that longer options are more often than not the correct answer. Naturally, it is impossible to make all options exactly the same length, but try to avoid situations where correct answers are more than one-and-a-half times the length of incorrect options. After eliminating the redundancies in the options and condensing the correct option, we have:

When 53 Americans were held hostage in Iran, the United States

- a. did nothing to try to free them.
- b. declared war on Iran.
- c. undertook diplomatic and military efforts to free them.
- d. expelled all Iranian students.

Item 8 has some problems, too. First, let's consider the use of "all of the above" and "none of the above." In general, "none of the above" should be used sparingly. Some item writers tend to use "none of the above" only when there is no clearly correct option presented. Students, however, can quickly catch on to such a practice and guess that "none of the above" is the correct answer without knowledge of the content being measured.

As far as "all of the above" goes, we cannot think of any circumstances in which its use may be justified. We recommend avoiding this option entirely.

The use of "both a and c" and "both b and c" was already discussed in relation to item 6. In that item, their use was appropriate and justifiable, but here it is questionable.

Again, let us see just what it takes to arrive at the correct choice, option e. Presumably, the item is intended to measure knowledge of square roots. However, the correct answer can be arrived at without considering square roots at all! A logical approach to this item, which would pay off with the right answer for someone who doesn't know the answer, might go something like this:

Sure wish I'd studied the square root table. Oh well, there's more than one way to get to the root of the problem. Let's see, 14 might be right, 16 might be right, and 4×4 might be right. Hmm, both a and c? No, that can't be it because I know that $4 \times 4 = 16$ and not 14. Well, both b and c have to be it! I know it's not "none of the above" because the teacher never uses "none of the above" as the right answer when she uses "both a and c" and "both b and c" as options.

When using "both a and c" and "both b and c," be on the alert for logical inconsistencies that can be used to eliminate options. Naturally, this problem can be minimized by using such options sparingly. Also try to monitor your item construction patterns to make sure you're not overusing certain types of options.

Finally, we come to a good item. Item 9 is free of the flaws and faults we've pointed out in this section. There are a lot of things to consider when you write test items, and keeping them all in mind will help you write better multiple-choice questions. But it's virtually impossible for anyone to write good items all the time. So, when you've written a poor item, don't be too critical of yourself. Analyze it, revise or replace it, and learn from your mistakes.

me

gan

2

nofer, Middlebury College

Nancy Van Note Chism, Indiana University—Purdue University, Indiana

ing Zhu and Matthew Kaplan, University of Michigan

polo, University of Michigan

edge, The Open University

in, University of Texas at Austin

University

at Austin

test once in this way will suggest the answers to questions that might have been difficult had they been answered in serial order. When you have gone through the test once in this fashion, go through it again and answer any questions that are now obvious. There will still usually remain a few questions that have been left unanswered. It is in connection with these that certain tricks may be useful.

If the item is multiple choice, don't simply guess at this stage of the game. See whether or not it is possible to eliminate some of the choices as incorrect. In a four-choice multiple-choice item, the probability of getting the answer right by pure guesswork is one in four; if you can eliminate two of them, your chances are 50-50. So take advantage of the mathematics of the situation.

Once the examination has been answered completely, it is a good idea to go through the whole thing again to check your choices on the various items to make sure that they are the ones you still regard as correct and to make sure that you have made no clerical errors in recording them. In this connection, it is worthwhile to point out the common misconception that, when you change your answers, you usually change from right answers to wrong ones. As a matter of fact, Mueller and Wasser (1977) reviewed 18 studies demonstrating that most students gain more than they lose on changed answers.

HELPING STUDENTS BECOME TEST-WISE

Particularly in the case of multiple-choice examinations, I have found that a good morale builder is spending 15 minutes or so the day before the first test telling students how to take a test of this sort.

Some of the points that I make in such a lecture follow.

Taking Multiple-Choice Tests

The student taking a multiple-choice examination is essentially in the same position as a poker player. The object is to get into a position where you are betting on a sure thing. If this is impossible, at least make your bet on the choice where the odds are in your favor. In poker, you are obviously in the strongest position if you know exactly what the opponent has; and in the examination situation, you are also in the strongest position if you know the material. There is no substitute for study. At the same time, it is unlikely that you will be absolutely certain of all the right answers. In these cases certain techniques may help.

What I recommend (to the student) is this: go through the examination a first time and answer all of the items you know. In addition to getting a certain amount of the examination done without wasting too much time on single, difficult items, it is frequently true that going through the complete

ADMINISTERING THE TEST

Handing out a test should be a simple matter. Usually it is, but in large classes, simple administrative matters can become disasters. It is hard to imagine how angry and upset students can become while waiting only ten minutes for the proctors to finish distributing the test forms. And if this doesn't move you, imagine your feelings when you find that you don't have enough tests for all of the students. (It has happened to me twice—deserving a place among my worst moments in teaching!)

How can you avoid such problems?

1. If you are having tests duplicated, ask for at least 10 percent extra—more if the test is administered in several rooms. (Some proctor always walks off with too many.) This gives you insurance against miscounting and against omitted or blank pages on some copies.
2. Unless there is some compelling reason to distribute the tests later, have your proctors pass out the tests as students come into the room. This protects students from mounting waves of panic while they wait for the tests to be distributed.
3. Minimize interruptions. Tell students before the exam that you will write announcements, instructions, or corrections on the blackboard. Some exam periods are less a measure of achievement than a test of the students' ability to work despite the instructor's interruptions.