CHAPTER **11**

Writing a Traditional Test

Some Valuable Ideas You'll Find in This Chapter

- Multiple-choice and matching questions can assess some kinds of thinking skills, including the abilities to apply and analyze.
- Well-designed multiple-choice questions give useful diagnostic information on where the student's thinking went wrong.
- Test blueprints—outlines of the learning goals covered by the test—are vital to planning effective multiple-choice tests.
- Avoid using "all of the above" and "none of the above" as multiple-choice options.
- Make multiple-choice tests challenging by making them open-book and open-note.

While performance assessments (Chapter Two) are growing in popularity, multiple-choice and other objective tests often still have a place in the assessment toolbox. This chapter discusses how to plan an objective test, how to write effective multiple-choice questions and other objective test items, and how objective tests can be used to assess some thinking skills. Essay test questions follow the principles for prompts (assignments) that are discussed in Chapter Ten.

What Is an Objective Test?

Objective tests might be defined as those that can be scored by a competent eight year old armed with an answer key, while subjective tests and assignments require professional judgment to score.

The most common kind of objective test item is the multiple-choice item. It consists of two parts. The *stem* asks the question; it may be phrased as a question or an incomplete sentence. The remainder of the test item is a set of *responses* or *options* from which the student chooses one answer. The incorrect options are called *distracters* or *foils* because their purpose is to distract or foil students who don't know the correct answer from choosing it.

Other kinds of objective test items are simply variations on the multiple-choice theme. A *true-false* item is a multiple-choice item with only two options. *Matching* items are a set of multiple-choice items with identical options. And a *completion* or *fill-in-the-blank* item is a multiple-choice item with no options.

Why Use an Objective Test?

Despite years of often-justified criticisms, objective tests remain widely used for the reasons noted in Table 2.8. They are especially good for assessing fundamental knowledge and understanding. They can also assess some important thinking skills, such as the ability to identify correct applications, examples, functions, causes, or effects. Interpretive exercises, a special kind of objective item discussed later in this chapter, can assess application, problem-solving, and analysis skills.

Well-constructed multiple-choice tests can also help diagnose problem areas. Consider this simple example:

What is 2 × .10?

 A. 20
 B. 2.10
 C. 0.2
 D. 0.02

Each distracter gives a clue as to where the student's thinking goes wrong. Choosing B, for example, indicates that the student confuses multiplication and addition signs. Choosing A indicates that the student confuses multiplication with division, and choosing D indicates a problem with decimal place value. Faculty and staff can thus use the results of well-written multiple-choice tests to identify areas of difficulty and help their students accordingly.

Objective tests are not always an appropriate assessment choice, however, for the reasons noted in Table 2.7. They cannot measure many important thinking skills and do not measure real-world performance as well as performance assessments (Chapter Two) do. They have other shortcomings as well:

- As you will learn in this chapter, writing good, clear items with good distracters can be difficult and time-consuming.
- Most objective tests require significant reading ability, so students who fail, say, a multiple-choice science test might truly understand the tested scientific concepts but fail because they read unusually slowly. While reading skill is essential for success in academic efforts, if we are assessing understanding of science concepts, we want an assessment tool that assesses just that.
- It's possible to guess the correct answer to most objective items through plain luck or through skill in being test-wise. While a well-constructed test minimizes test-wise students' advantage, there is always the possibility that a student who doesn't know the material will do well on a test through chance.

Planning an Objective Test with a Test Blueprint _____

Effective assignments are planned by developing a rubric: a list of the learning goals that students are to demonstrate in the completed assignment (Chapter Nine). Effective tests are similarly planned by developing a *test blueprint*: an outline of the test that lists the learning goals that students are to demonstrate on the test. Test blueprints are critical to creating effective tests for several reasons:

Test blueprints help ensure that the test focuses on the learning goals you think are most important.   Suppose that you are writing a test for units 8, 9, and 10 of a particular course. While you consider unit 10 the most important of the three, you may find that it's much easier to think of test questions for unit 8. If you write your test without a blueprint, you can easily end up with too many questions on unit 8 and too few on unit 10. Students taking such a test may be able to do fairly well without having mastered important concepts of unit 10.

Test blueprints help ensure that a test gives appropriate emphasis to thinking skills.   Faculty writing test questions without a blueprint often find that questions asking for simple conceptual knowledge are easier to write than those asking students to interpret, infer, analyze, or think in other ways. Tests written without blueprints can thus become tests of trivia rather than of thinking skills. Students who do well on such tests may not have mastered important skills, and students who have truly learned those important skills may earn low scores.

Test blueprints make writing test questions much easier. Armed with a test blueprint, you'll know exactly what must be covered on the test (one question on concept A, two on skill B, and so on), and you'll spend less time pondering what questions to write.

Test blueprints help document that students have achieved major learning goals. Test scores or grades alone give little direct information on exactly what students have learned (Chapter One). But if those scores or grades are accompanied by test blueprints that describe the learning goals covered on the tests, you'll have direct evidence of exactly what students who did well on the tests have learned. This makes test blueprints an important part of an assessment program. Exhibit 16.3 shows how test results can be mapped back to the test blueprint.

Begin creating a test blueprint by using syllabi, lecture notes, readings, and other curricular materials to list the major areas that the test will cover. A midterm exam, for example, might cover the first five chapters in the textbook. A comprehensive exam for graduating seniors might cover the six courses required of all majors.

Next, allocate fractions of the test to each of those areas, in proportion to the relative importance of each area, by assigning points or a number of test questions to each area (perhaps twenty points or four questions to unit 1, thirty points or six questions to unit 2, and so on).

Next, within each area, list the learning goals you want to assess. To make clear to both yourself and your students exactly what you want them to demonstrate on thẻ test, phrase the learning goals using action verbs that describe what students should be able to think and do (Chapter Eight). Instead of simply listing "Hemingway" in a test blueprint on twentieth-century American literature, state the knowledge and skills you want students to demonstrate regarding Hemingway ("Identify works written by Hemingway," "Distinguish Hemingway's writing style from those of his peers").

Finally, spread the points or test questions within each area among the learning goals within that area, again in proportion to their importance. For example, if you have allocated four questions to unit 1 and list four learning goals within that chapter, you may decide to have one question on each learning goal. But perhaps one of the learning goals is especially important. In that case, you may decide to have two questions on that goal, and you will need to eliminate one of the other unit 1 learning goals from the test.

Exhibit 11.1 is an example of a test blueprint. It has one test question on each topic unless otherwise indicated.

Exhibit 11.1. A Test Blueprint for an Examination in an Educational Research Methods Course

Sampling

- Understand the difference between a sample and a population.
- Understand how each type of sample is selected.
- Understand how to choose an appropriate sample size.

Instrumentation and Survey Research

- Understand the relative merits and limitations of published and locally developed instruments.
- Recognize examples of each of the four frames of reference for interpreting scores.
- Understand appropriate uses of each item format (such as multiple choice and Likert scale).
- Understand the characteristics of a good instrument item, including how to avoid biased questions.

Descriptive Statistics

- Select the most appropriate descriptive statistic for a given research situation.
- Use percentage guidelines to interpret standard deviations.
- Identify the direction and strength of r and/or a scatter plot.

Validity and Reliability

- Identify the type of reliability or validity evidence provided by given information on an instrument. (two questions)
- Understand the meaning and implications of measurement error.
- Recognize examples of measurement error in a given situation.
- Understand the general principles for ensuring validity.

Inferential Statistics

- Select the most appropriate inferential statistic (t, F, or χ^2) for a given research situation. (two questions)
- Know the most common cutoff points that statisticians use in deciding whether two means differ statistically significantly from one another.
- Interpret the results of t-tests as presented in research articles.

Experimental Research

- Interpret the symbolic representations of experimental designs.
- Identify the appropriate research design for a given research situation.

Correlational Research

- Understand what r^2, R, and R^2, and partial correlations are and what they tell us. (two questions)
- Understand what multiple regression analysis is used for and what it tells us.

Once you have finalized the test blueprint, give copies to your students to help them focus their studies on the learning goals that you think are most important.

Writing Good Multiple-Choice Items _____

As with any other assessment, multiple-choice tests should yield fair and truthful information on what students have learned (Chapter Three). There are just two basic precepts to writing fair and truthful multiple-choice items.

- *Remove all the barriers that will keep a knowledgeable student from answering the item correctly.* Students who have truly learned the concept or skill that a particular item tests should choose the correct answer.
- *Remove all clues that will help a less-than-knowledgeable student answer the item correctly.* Students who truly haven't learned the concept or skill that a particular item tests should answer the item incorrectly.

The suggestions in Table 11.1 follow these two precepts. Exhibit 11.2 gives some examples of multiple-choice items that follow most of these suggestions and assess thinking skills as well as conceptual understanding.

Writing good multiple-choice items can be difficult. Test publishers write, try out, and discard many, many items for each one that ends up in a published test. Even the examples in Exhibit 11.2 don't

Table 11.1. Tips for Writing Good Multiple-Choice Questions

General Tips

Keep each item as concise as possible. Short, straightforward items are usually easier to understand than complex statements. Avoid irrelevant material, digressions, and qualifying information unless you are specifically assessing the skill of identifying needed information. Don't repeat the same words over and over in the options; put them in the stem.

Define all terms carefully. If you ask, "Which of the following birds is largest?" make clear whether you mean largest in terms of wingspan or weight. What do you mean by "sometimes," "usually," or "regularly"?

Don't make the vocabulary unnecessarily difficult. Except for terms you are specifically assessing, keep the vocabulary simple—perhaps high school level. Otherwise you may unfairly penalize students who know the material but don't have a strong general vocabulary.

Watch out for "interlocking" items: items in which a student can discern the answer to one question from the content of another. Review carefully all items that share similar options. In a similar vein, don't ask students to use their answer to one question to answer another. If they get the first question wrong, they will automatically get the other question wrong as well, even if they understand the concept tested in the second question.

Writing a Good Stem

The stem should ask a complete question. The student shouldn't have to read the options to discern the question. To check this, ask yourself if students would be able to answer the question posed in the stem correctly if no options were provided.

Avoid "Which of the following" items. They require students to read every option and can penalize slow readers in a timed-testing situation.

Don't ask questions that can be answered from common knowledge. Someone who hasn't studied the material shouldn't be able to answer the questions correctly.

Avoid negative items. In a stressful testing situation, students can miss the word *not* or *no*. If you must have negative items, underline, capitalize, or boldface words like NOT or EXCEPT.

Avoid grammatical clues to the correct answer. Test-wise students know that grammatically incorrect options are wrong. Use expressions like "a/an," "is/are," or "cause(s)."

Writing Good Options

You needn't have the same number of options for every question. Four options are fine. A good fifth option is often hard to come up with, takes extra reading time, and reduces the chances of guessing the correct answer only from 25 to 20 percent. Some questions may have only three plausible options (for example, "Increases," "Decreases," and "Remains unchanged").

Order responses logically. Order responses numerically if they are numbers and alphabetically if they are single words. This helps students who know the answer find it quickly. If the options have no intuitive order, insert the correct answer into the responses randomly.

Line up responses vertically rather than horizontally. It's much easier—and less confusing—to scan down a column than across a line to find the correct answer. If you are using a paper test and your options are so short that this seems to waste paper, arrange the test in two columns.

Make all options roughly the same length. Test-wise students know that the longest option is often the properly qualified, correct one. (For this reason, a relatively long option can make a good distracter!)

Avoid repeating words between the stem and the correct response. Test-wise students will pick up this clue. (On the other hand, verbal associations between the stem and a distracter can create an effective distracter.)

Avoid using "None of the above." A student may correctly recognize wrong answers without knowing the right answer. Use this option only when it is important that the student know what *not* to do. If you use "none of these," use it in more than one question, both as a correct answer and an incorrect answer.

Avoid using "All of the above." This option requires students to read every option, penalizing those in a timed testing situation who know the material but are slow readers. Students who recognize option A as correct and choose it without reading further are also penalized. "All of the above" also gives full credit for incomplete understanding; some students may recognize options A and B as correct and therefore correctly choose "All of the above" even though they don't recognize option C as correct.

Writing Good Distracters

The best distracters help diagnose where each student went wrong in his or her thinking. Identify each mental task that students need to do to answer a question correctly, and create a distracter for the answer students would arrive at if they completed each step incorrectly.

Use intrinsically true or at least plausible statements. Test-wise students recognize ridiculous statements as wrong. To see if your test has such statements, ask a friend who has never studied the subject to take the test. His or her score should be roughly what would be earned from guessing randomly on every item (25 percent for a four-option multiple-choice test).

Exhibit 11.2. Multiple-Choice Questions on Assessment Concepts

Correct answers are in italics.

1. Which statement refers to measurement as opposed to evaluation?
   A. *Emily got 90 percent correct on the math test.*
   B. Lin's test scores have increased satisfactorily this year.
   C. Justin's score of 20 on this test indicates that his study habits are ineffective.
   D. Keesha got straight A's in her history courses this year.

2. Alyssa took a test on Tuesday after a big fight with her parents Monday night. She scored a 72. Her professor let her retake the same test on Thursday when things cooled off. She scored 75. The difference in her scores may be attributed to:
   A. chance or luck.
   B. lack of discrimination.
   C. lack of validity.
   D. *measurement error.*

3. People who score high on the Meyers Musical Aptitude Scale usually score low on the Briggs Biologists Aptitude Test. People who score low on the Meyers usually score high on the Briggs. Which of the figures below *most likely* represents the correlation between the two tests?
   A. .80
   B. .00
   C. −.10
   D. *−.60*

4. Choose the *most likely* correct answer to this nonsense question, based on what you know about informed guessing on tests. A drabble will coagulate under what circumstances?
   A. Only when pics increase
   B. Only when pics change color
   C. By drawing itself into a circle
   D. *Usually when pics increase, but occasionally when pics decrease*

---

completely follow the suggestions in Table 11.1. So don't expect to be able to follow all these suggestions all the time, and don't expect your test questions to work perfectly the first time you use them. Analyze the results (Chapter Sixteen), revise the test accordingly, and within just a few cycles you will have a truly good test.

Avoiding Trick Questions

Many tests aim to include a few difficult questions to challenge the very best students. Unfortunately, writing difficult multiple-choice questions is, well, difficult. It's very hard to come up with meaningful distracters that will foil all but the best students.

Table 11.2. Tips for Writing Challenging Rather Than Trick Questions

Use a test blueprint. It ensures that each item assesses an important concept or skill.

Make your tests open-book, open-note. Tell students they can bring to the test anything they like except a friend or the means to communicate with one. Using open-book, open-note tests forces you to eliminate items assessing simple knowledge that students can look up. Your test will include only items that assess deeper comprehension and thinking skills.

Build items around common misconceptions. Many people, for example, think that plants get nutrients only from soil and water, not air; this misconception can become the basis of an effective botany test question.

Create interpretive exercises (discussed below). They assess thinking skills such as application and analysis.

Evaluate your test results using the tools in Chapter Sixteen. Revise any misleading or unnecessarily difficult items before including them in another test.

Too often, difficult questions are trick questions that focus on trivia or some finely nuanced point rather than an important concept. Follow the suggestions in Table 11.2 to write meaningful, challenging multiple-choice questions that assess important learning goals rather than trick questions.

Some faculty pride themselves on writing complex multiple-choice questions that require well-developed analytical reasoning skills to understand and answer correctly. These kinds of question can be fine *if*:

- One of the learning goals of the course is to develop these skills; *and*

- Students have opportunities to learn and practice these skills before taking the test; *and*

- This learning goal is reflected in the test blueprint.

Interpretive Exercises

Interpretive exercises, sometimes referred to as context-dependent or enhanced multiple-choice items or scenario testing, consist of a stimulus, such as a reading passage or a chart, that students haven't seen before, followed by a set of objective items. Many published aptitude and achievement tests include interpretive exercises.

Interpretive exercises have three defining characteristics:

- *The stimulus material must be new to the students;* they must never have seen it before. This requires students to apply what they have learned to a new situation, making interpretive exercises a good way to assess application skills.

- *Students must read or examine the material in order to answer the objective items that follow.* They should not be able to answer

any of the items simply from their general understanding of what they've learned in the course or program. Again, this defining characteristic makes an interpretive exercise an assessment of application skill rather than simple conceptual understanding.

- *The items must be objective, with one and only one correct answer for each item.* If you ask students to write or otherwise create something in response to the stimulus, you have a performance assessment (Chapter Two), not an interpretive exercise.

While all interpretive exercises by definition assess skill in applying knowledge and skills to new situations, interpretive exercises can also assess skill in generalizing, inferring, concluding, problem solving, and analysis. Performance assessments can assess these skills as well, but interpretive exercises are more efficient. Given the same amount of work time, students will give you information on a broader range of skills through interpretive exercises than they can through performance assessments, and you can score interpretive exercises more quickly.

Interpretive exercises are not always appropriate, however. As with other objective items, they can be difficult and time-consuming to write. If the stimuli are reading passages, interpretive exercises may unfairly penalize students who have the knowledge and skills being assessed but are slow readers. And although interpretive exercises are good for assessing some thinking skills, they cannot assess some other thinking skills such as organizing, defining problems, and creating.

Table 11.3. Tips for Writing Good Interpretive Exercises

Keep the size of the stimulus in proportion to the questions asked. Having students read a full page of text in order to answer only three questions is hardly an efficient use of their time. Generally aim to ask at least three questions about any stimulus, and ask more about longer stimuli.

Be on the lookout for interlocking items. They seem to crop up more often in this format.

Give students realistic scenarios. Possibilities include:

A chart, diagram, map, or drawing with real or hypothetical information

A brief statement written by a scholar, researcher, or other significant individual

A passage from a novel, short story, or poem

A description of a real or imaginary scenario, such as a scientific experiment or a business situation

For foreign language courses and programs, any of the above written in the language

Be creative! Remember that the stimulus need not be a reading passage; it can be a diagram, or picture, or chart, as shown in Exhibit 11.3.

The key to writing good interpretive exercises is to keep in mind the three defining characteristics listed above. Table 11.3 offers other suggestions for creating effective interpretive exercises.

Exhibit 11.3 is an example of an interpretive exercise that follows most of these suggestions. (Correct answers are in italics. Item analysis—the concept assessed by these items—is discussed in Chapter Sixteen.)

Exhibit 11.3. An Example of an Interpretive Exercise

Items 1 to 5 refer to the item analysis information given below. The correct options are marked with an asterisk.

Item 1	A	B*	C	D
Top third			10	
Bottom third	1	4	3	2

Item 2	A*	B	C	D
Top third	8		2	
Bottom third		7	3	

Item 3	A	B	C*	D
Top third	5		1	4
Bottom third	2		4	4

Item 4	A*	B	C	D
Top third	10			
Bottom third	9	1		

Write the item number (1, 2, 3, or 4) in the space provided.

1. _4_ Which item is easiest?
2. _3_ Which item shows negative (*very* bad) discrimination?
3. _2_ Which item discriminates *best* between high and low scores?

For the remaining items, write the option letter (A, B, C, or D) in the space provided.

4. _B_ In item 2, which distracter is most effective?
5. _B_ In item 3, which distracter *must* be changed?

Matching Items

Matching items are a set of multiple-choice items with a common set of responses. If, as you write a multiple-choice test, you find yourself writing several items with similar options, consider converting them into a matching set.

Matching items are even more efficient than multiple-choice questions. Because students need to read only one set of options to answer several items, they can often answer five well-written matching items more quickly than five multiple-choice items, giving you more assessment information in a given amount of testing time. They are also faster to write because you don't have to come up with a fresh set of distracters for each item.

Matching items are a good way to assess certain kinds of basic knowledge. Students can match terms and definitions, causes and effects, authors and titles of their works, people and their achievements, foreign words and their English translations, or tools and their uses. Matching items need not be entirely verbal; students can match symbols with the concepts they represent, pictures of objects with their names, or labeled parts of a pictured entity (say, a microscope or a cell) with their functions.

Matching items can also assess some thinking skills, especially the ability to apply what students have learned to new situations and the ability to analyze interrelationships. Students can match concepts with new examples of them, causes with likely effects, and hypothetical problems with the tools, concepts, or approaches needed to solve them.

Table 11.4 offers suggestions for creating good sets of matching items—possibly quite different from matching items you may have used or seen in the past.

Exhibit 11.4 gives an example of a matching set that follows these suggestions and assesses application skill as well as conceptual understanding.

True-False Items

True-false items are multiple-choice items with only two options. Their most common use is simple knowledge: Is a given statement correct or not? But they can also be used in other situations with just two possible answers: Is a statement fact or opinion? Is it an example of qualitative or quantitative assessment evidence? Does A cause B?

True-false items have no particular advantages beyond those of any other objective format. Because they have the following serious shortcomings, they should be used rarely, if at all:

Table 11.4. Tips for Writing Good Matching Items

A matching set should consist of homogeneous items. Every option in the answer key should be a plausible answer for every item or question. Otherwise test-wise students will quickly eliminate implausible answers, and students who are not test-wise will read and consider the full set of responses over and over.

Make an imperfect match between the two columns. Allow students to use each option more than once or not at all. A perfect match (in which each option is the answer for exactly one item) gives an unfair advantage to test-wise students, who will cross out each option as it's chosen and then guess among the remaining options. A perfect match also gives an unfair disadvantage to students who misunderstand one item but truly know all the other answers; if they choose one incorrect answer, they must, by process of elimination, choose a second incorrect answer, because it's the only option left.

Make it easy for students who know the material to find the correct answer. Make the longer statements the "questions," and limit the answer key to single words or short phrases. For example, list definitions as the questions and the terms they define as the options. Or list accomplishments as the questions and the people achieving them as the options. Otherwise students will need to keep scanning through a list of lengthy options to find the correct answers. This penalizes those who have learned the material but are slow readers. Limit the number of matching items in a set to no more than ten or so, and keep the entire exercise on one page. Arrange the options in a logical order (usually alphabetically).

Give clear directions. In an introductory sentence, explain how the two columns are related—for example, "Match each theory with the person who conceived it." Point out that options may be used more than once or not at all; your students may never have seen this kind of matching set before. Give each column an explanatory title (for example, "Theory" and "Originator") if that would be helpful.

Be inventive! The answer key need not be words or phrases in a column; it can be lettered parts of a diagram, drawing, map, or chart.

Exhibit 11.4. Matching Items from a Nursing Research Methods Test

In this set of matching items, some options may be used more than once or not at all. Correct answers are in italics.

Match each measurement with its level of measurement.

    A. Interval

    B. Ordinal

    C. Nominal

    D. Ratio

D 1. Fluid intake, in ounces, of a postsurgical patient

C 2. Religious affiliation

D 3. Medication dosage

C 4. Type of adjuvant therapy (chemotherapy, hormonal therapy, or radiation therapy)

B 5. Level of patient advocate support for a patient (very supportive, moderately supportive, somewhat supportive, not supportive)

Source: Adapted with permission from test questions written by Christina Barrick, associate professor of Nursing, Towson University.

Table 11.5. Tips for Writing Good True-False Items

Keep them simple. Avoid lengthy qualifiers and broad generalizations, which can be confusing and hard to make plausible as true or false statements.

Use them only to assess important learning goals. It's easy for true-false items to descend into trivia.

Avoid negative and double-negative statements. These are confusing in a true-false format.

Keep the proportion of true statements close to but not exactly 50 percent. Test-wise students will scan the number of true statements they've marked and use that to decide how to guess on the items they don't know.

- Students who haven't learned the material have a high probability (50 percent) of guessing the correct answer.
- Unlike multiple-choice and matching items, true-false items give no clues about where students who answered incorrectly went wrong in their thinking.
- It's difficult to write true-false items assessing thinking skills, although they can be used in interpretive exercises.
- For classic true-false items—those giving true or false statements—it can be very difficult to write unambiguous, unqualified statements that are either definitely true or definitely false.
- Students may correctly recognize a false statement without knowing its true counterpart.

If true-false items appear to be appropriate for your situation, the suggestions in Table 11.5 will help make the best of them.

Completion or Fill-in-the-Blank Items

Completion items are multiple-choice items with no options provided. They pose questions that students answer with a word, number, or symbol. A fill-in-the-blank item is a completion item posed as a sentence with a missing word, number, or symbol.

To be true objective items, completion items should have only one correct answer. Recall the definition of an objective test given at the beginning of this chapter: a test is objective if a reasonably competent eight year old armed with an answer key can score it accurately. True completion items can be scored in this fashion. Many short-answer items are really subjective, with a number of acceptable answers that require professional judgment to score. While such subjective items may be an appropriate part of an assessment program, considering the time they take to score and the limited information they provide, performance assessments (Chapter Two)

Table 11.6. Tips for Writing Good Completion or Fill-in-the-Blank Items

---

Keep all blanks or spaces for recording answers of uniform length. Blanks or spaces of varying length give test-wise students clues. To facilitate scoring, have students record all their answers in a column on one side of the page. If you are using fill-in-the-blank items, make the blanks in the sentences very short placeholders, and have students write their answers in a column of longer blanks.

If you are using fill-in-the-blank items, structure sentences so the blanks are toward the ends of the sentences. Sentences will be easier for your students to understand than if the blanks are at the beginning.

Avoid lifting sentences out of a textbook. Too often the resulting items are ambiguous or focus on trivia.

---

may be a better choice, as they assess thinking skills and give students more opportunities to acquire and demonstrate deep, lasting learning.

Completion items are a good choice for assessing those essential facts that must be memorized and should not be guessed from multiple-choice items. They are also appropriate when the correct answer would be easy for students to recognize in a multiple-choice format. Completion items are widely used in mathematics, for example, when a test-wise student might deduce the correct multiple-choice answer by working backward from each option. They can be a good way to develop multiple-choice distracters for future tests: simply choose the most common incorrect answers as foils.

Truly objective completion items rarely assess thinking skills except in mathematics. Because scoring is difficult to automate, this format is not a good choice for large-scale assessment programs.

The key to writing truly objective completion items is to design them so that one specific word, number, or symbol is the only correct answer. Table 11.6 offers other suggestions.

Pulling an Objective Test Together

Before assembling items into a test, review them in terms of these questions:

- *Do the items follow the test blueprint?* Does each assess an important learning outcome, or are any of them trick questions that ask about trivia? Are the formats and content appropriate for the learning goals you are assessing?

- *Are the items at an appropriate reading level?* Other than vocabulary terms that you are specifically assessing, are the items

simple, clear, and straightforward? Are they free of excessive verbiage?

- *Would experts agree on the answers?*
- *Are the items of appropriate difficulty?* Ideally, there should be a few easy items on fundamental concepts that virtually everyone answers correctly. Perhaps there are also a few items designed to challenge the top students. An excessively difficult test, no matter how you curve the scores, is frustrating and demoralizing for students. See Chapter Sixteen for further discussion of item difficulty.
- *Are there any interlocking items* or items with any other clues for test-wise students?

Next, order the items. The first ones should be the easiest, to reassure the test anxious, and quickly answered, to help those who aren't test-wise. The last items should be the most difficult and the most complex (requiring the most thinking time). Interpretive exercises often go toward the end.

Then write directions that explain:

- The purpose of the test
- How the answers will be scored
- How to answer (Can they choose more than one answer? Should they choose the one best answer? Is guessing encouraged?)
- How to record answers
- Any time limits (if the test is lengthy and timed—a two-hour final exam, for example—you may want to suggest time limits for each section)

Finally, let the test sit for twenty-four hours after you have completed it, and then proofread it one last time. Prepare the scoring key before the test is duplicated or posted online, as the process of preparing the key can identify typos and unclear items missed in earlier readings.

Should Students Explain Their Answers to Objective Items?

A major concern with objective items, especially true-false items, is the possibility that students who haven't learned the material can still guess the correct answer. One way to solve this problem is to ask students to write brief explanations of why they chose their answer. For true-false items, students can be asked to correct any statements they mark as false.

While this does eliminate the possibility of students' blind guessing the correct answer, it also removes one of the fundamental advantages of objective items: their efficiency. Students won't be able to answer as many questions in a given amount of testing time, so the test will assess a narrower range of goals. And rather than having a scanner or a competent eight year old score the tests, you must read every answer and use your professional judgment to decide which are correct and which are not, which takes far more of your time and energy.

If you want students to explain their answers, consider instead giving them an assignment or essay question (Chapter Ten) for which they must compose a more complete written response. This will elicit deeper thought, give you richer assessment information, and give your students a better learning experience.

Should Students Be Encouraged to Guess?

In a word, yes. The "correction for guessing" used on some published tests aims to equate the scores of students who answer questions randomly with those who leave questions blank. Test-wise students readily guess on items of which they're unsure. They know that if they can eliminate even one option as implausible, they raise their odds of guessing correctly beyond random chance. If students who aren't test-wise aren't encouraged to guess, they're being unfairly penalized for not being sufficiently test-savvy.

Time to Think, Discuss, and Practice _____

Write each of the following for a unit or concept that you teach or have studied, following as many of the guidelines in this chapter as you can. Share your drafts with group members for comments and suggestions.

1. Six multiple-choice items
2. A set of matching items
3. An interpretive exercise

Recommended Readings _____

The following readings are recommended along with the references cited in this chapter.

Berk, R. (1998). A humorous account of 10 multiple-choice test item flaws that clue testwise students. *Journal on Excellence in College Teaching, 9*(2), 93–117.

Clegg, V. L., & Cashin, W. E. (1995). *Improving multiple choice tests* (IDEA Paper No. 16). Manhattan: Kansas State University, Center for Faculty Evaluation and Development.

Frary, R. B. (1995). More multiple-choice item writing do's and don'ts. *Practical Assessment, Research and Evaluation, 4*(11). Retrieved September 16, 2008, from http://PAREonline.net/getvn.asp?v=4&n=11.

Gronlund, N. E. (2005). *Assessment of student achievement* (8th ed.). Needham Heights, MA: Allyn & Bacon.

Haladyna, T. M. (1997). *Writing test items to evaluate higher order thinking*. Needham Heights, MA: Allyn & Bacon.

Haladyna, T. M. (2004). *Developing and validating multiple-choice test items* (3rd ed.). Needham Heights, MA: Allyn & Bacon.

Kehoe, J. (1995). Writing multiple-choice items. *Practical Assessment, Research and Evaluation, 4*(9). Retrieved September 16, 2008, from http://PAREonline. net/getvn.asp?v=4&n=9.

Kubiszyn, T., & Borich, G. D. (2002). *Educational testing and measurement: Classroom application and management* (7th ed.). San Francisco: Jossey-Bass.

Linn, R. L., & Gronlund, N. E. (2000). *Measurement and assessment in teaching* (8th ed.). New York: Macmillan.

CHAPTER 7

WRITING OBJECTIVE TEST ITEMS

THUS FAR we have discussed how to establish general goals for instruction and how to develop instructional objectives derived from those goals. We have also discussed using the test blueprint to ensure an adequate sampling of the content area, accurate matching of test items to instructional objectives, and that our test measures higher-level thinking skills. We are now ready to put some "meat" on this test "skeleton." In this chapter we will discuss objective test items and how to construct them. Objective test items include items with the following formats: true–false, matching, multiple choice, and completion or short answer. The essay item, because of its special characteristics and scoring considerations, will be treated in Chapter 7, along with methods to measure the way students organize knowledge and suggestions for open-book exams.

WHICH FORMAT?

Once you have reached the item-writing stage of test construction, you will have to choose a format or combination of formats to use for your test. Although your choice can be somewhat arbitrary at times, this is not always the case. Often your decision has already been made for you, or, more correctly, you may have at least partially made the decision when you wrote the objective or objectives. In many instances, however, you will have a choice among several item formats. For example, consider the following objectives and item formats.

OBJECTIVE 1: Given a story, the student can recognize the meaning of all new words.

TEST ITEMS: Circle the letter that represents the correct meaning of the following words:

1. intention
   a. desire
   b. need
   c. direction
   d. command

2. crisis
   a. feeling
   b. message

c. pending

d. critical

OBJECTIVE 2: The student can associate the characteristics of leading characters with their names.

TEST ITEM: The first column is a list of the names of the main characters in *Huckleberry Finn*. Descriptions of the main characters are listed in the second column. In the space provided, write the letter of the description that matches each character.

Character	*Description*
_____ **1.** Tom	**a.** Cruel
_____ **2.** Becky	**b.** Always by himself, a loner
_____ **3.** Jim	**c.** Popular, outgoing, fun loving
_____ **4.** Huck	**d.** Always critical
_____ **5.** Mrs. Watson	**e.** Sneaky, lying, scheming
	f. Kind, gentle, loving
	g. Dull, slow moving

OBJECTIVE 3: The student can write a plausible alternative ending to a story.

TEST ITEM: You have just read the story *Huckleberry Finn*. In 40 words, write a different ending to the story that would be believable.

OBJECTIVE 4: The students will recognize whether certain events occurred.

TEST ITEM: Here is a list of incidents in *Huckleberry Finn*. Circle T if it happened in the story and F if it did not.

1. The thieves were killed in the storm on the river.   T   F

2. Jim gained his freedom.   T   F

3. Tom broke his leg.   T   F

Objective 1 was measured using a multiple-choice format. We might also have tested this objective using a true–false or matching format. Similarly, objective 2 lends itself to a multiple-choice as well as a matching format. Since in many circumstances alternative item formats may be appropriate, the choice between them will be made on the basis of other considerations. For example, time constraints or your preference for, or skill in, writing different types of items will undoubtedly influence your choice of item format.

Objective 3, however, requires an essay item. There is no way this objective can be measured with an objective item. Objective 4 lends itself almost exclusively to a true–false format. Perhaps other formats would work, but the true–false format certainly does the job. In short, there are times our objectives tell us which format to use. At other times we must consider other factors. Let's look more closely at the different item formats.

TRUE–FALSE ITEMS

True–false items are popular probably because they are quick and easy to write, or at least they seem to be. Actually, true–false items do take less time to write than good objective items of any other format, but good true–false items are not that easy to write. Consider the following true–false items. Use your common sense to help you determine which are good items and which are poor.

> **EXERCISE:** *Put a G in the space next to the items you believe are good true–false items and a P next to the items you feel are poor.*
>
> _____ 1. High-IQ children always get high grades in school.
> _____ 2. Will Rogers said, "I never met a man I didn't like."
> _____ 3. If a plane crashed on the Mexican-U.S. border, half the survivors would be buried in Mexico and half in the United States.
> _____ 4. The use of double negatives is not an altogether undesirable characteristic of diplomats and academicians.
> _____ 5. Prayer should not be outlawed in schools.
> _____ 6. Of the objective items, true–false items are the least time-consuming to construct.
> _____ 7. The trend toward competency testing of high school graduates began in the late 1970s and represents a big step forward for slow learners.
>
> *Answers:* 1. P; 2. G; 3. P; 4. P; 5. P; 6. G; 7. P.

In item 1, the word *always* is an absolute. To some extent, true–false items depend on absolute judgments. However, statements or facts are seldom completely true or completely false. Thus, an alert student will usually answer "false" to items that include *always, all, never*, or *only*.

To avoid this problem, avoid using terms like *all, always, never*, or *only*. Item 1 could be improved by replacing always with a less absolute term, perhaps *tend*. Thus item 1 might read:

High-IQ children tend to get high grades in school.

Item 2 is a good one. To answer the item correctly, the students would have to know whether Will Rogers made the statement. Or do they? Consider the following situation:

Mrs. Allen, a history teacher and crusader against grade inflation, couldn't wait to spring her latest creation on her students. She had spent weeks inserting trick words, phrases, and complicated grammatical constructions into her 100 item true–false test. In order to ensure low grades on the test, she allowed only 30 minutes for the test. Although her harried students worked as quickly as they could, no one completed more than half the items, and no one answered more than 40 items correctly. No one, that is, except Tina. When Tina handed her test back to Mrs. Allen after two minutes, Mrs. Allen announced, "Class, Tina has handed in her test! Obviously, she hasn't read the questions and will earn a zero!" When she scored the test, however, Mrs. Allen was shocked to see that in fact Tina had answered 50 items correctly. She earned the highest score on the test without even reading Mrs. Allen's tricky questions. Confused and embarrassed, Mrs. Allen told the class they would have no more true–false tests and would have essay tests in the future.

This points to the most serious shortcoming of true–false items: With every true–false item, regardless of how well or poorly written, the student has a 50% chance of guessing correctly even without reading the item! In other words, on a 50-item true–false test, we would expect individuals who were totally unfamiliar with the content being tested to answer about 25 items correctly. However, this doesn't mean you should avoid true–false items entirely, since they are appropriate at times. Fortunately, there are ways of reducing the effects of guessing. Some of these are described next and another will be presented in Chapter 11.

1. Encourage all students to guess when they do not know the correct answer. Since it is virtually impossible to prevent certain students from guessing, encouraging all students to guess should equalize the effects of guessing. The test scores will then reflect a more or less equal "guessing factor" plus the actual level of each student's knowledge. This will also prevent test-wise students from having an unfair advantage over non-test-wise students.

2. Require revision of statements that are false. With this approach, space is provided for students to alter false items to make them true. Usually, the student also under-lines or circles the false part of the item. Item 1 is revised here along with other examples.

> T   F   High-IQ children *always* get high grades in school. *tend to*
>
> T   F   Panama is *north* of Cuba. *south*
>
> T   F   *September* has an extra day during leap year. *February*

With such a strategy, full credit is awarded only if the revision is correct. The disadvantage of such an approach is that more test time is required for the same number of items and scoring time is increased.

Item 3 is a poor item, but Mrs. Allen would probably like it because it is a trick question. "Survivors" of a plane crash are not buried! Chances are that you never even noticed the word *survivors* and probably assumed the item referred to fatalities. Trick items may have a place in tests of critical reading or visual discrimination (in which case they would no longer be trick questions), but seldom are they appropriate in the average classroom test. Rewritten, item 3 might read:

> If a plane crashes on the Mexican-U.S. border, half the fatalities would be buried in Mexico and half in the United States.

Item 4 is also poor. First of all, it includes a double negative—*not* and *undesirable*. Items with a single negative are confusing enough. Negating the first negative with a second wastes space and test-taking time and also confuses most students. If you want to say something, say it positively. The following revision makes this item slightly more palatable.

> The use of double negatives is an altogether desirable trait of diplomats and academi-cians.

We said slightly more palatable because the item is still troublesome. The word altogether is an absolute, and we now know we should avoid absolutes, since there usually are exceptions to the rules they imply. When we eliminate altogether the item reads:

> The use of double negatives is a desirable trait of diplomats and academicians.

However, the item is still flawed because it states an opinion, not a fact. Is the item true or false? The answer depends on whom you ask. To most of us, the use of double negatives is probably undesirable, for the reasons already stated. To some diplomats, the use of double negatives may seem highly desirable. In short, true–false statements should normally be used to measure knowledge of factual information. If you must use a true–false item to measure knowledge of an opinionated position or statement, state the referent (the person or group that made the statement or took the position), as illustrated in the following revision:

> According to the National Institute of Diplomacy, the use of double negatives is a desirable trait of diplomats and academicians.

Item 5 further illustrates this point. It is deficient because it states an opinion. It is neither obviously true nor obviously false. This revision includes a referent that makes it acceptable.

> The American Civil Liberties Union (ACLU) has taken the position that prayer should *not* be outlawed in schools.

Notice the word *not* in Item 5. When you include a negative in a test item, highlight it in italics, underlining, or uppercase letters so the reader will not overlook it. Remember that, unlike Mrs. Allen, you intend to determine whether your students have mastered your objective, not to ensure low test scores.

Item 6 represents a good item. It measures factual information, and the phrase "Of the objective items" qualifies the item and limits it to a specific frame of reference.

The last item is deficient because it is double barreled. It is actually two items in one. When do you mark true for a double-barreled item? When both parts of the item are true? When one part is true? Or only when the most important part is true? The point is that items should measure a single idea. Double-barreled items take too much time to read and comprehend. To avoid this problem, simply construct two items, as we have done here:

> The trend toward competency testing of high school graduates began in the late 1970s.

> The trend toward competency testing represents a big step forward for slow learners.

Better? Yes. Acceptable? Not quite. The second item is opinionated. According to whom is this statement true or false? Let's include a referent.

> According to the Office of Education, the trend toward competency testing of high school graduates is a big step forward for slow learners.

Whose position is being represented is now clear, and the item is straightforward.

Suggestions for Writing True–False Items

1. The desired method of marking true or false should be clearly explained before students begin the test.

2. Construct statements that are definitely true or definitely false, without additional qualifications. If opinion is used, attribute it to some source.

3. Use relatively short statements and eliminate extraneous material.

4. Keep true and false statements at approximately the same length, and be sure that there are approximately equal numbers of true and false items.

5. Avoid using double-negative statements. They take extra time to decipher and are difficult to interpret.

6. Avoid the following:
   a. verbal clues, absolutes, and complex sentences.
   b. broad general statements that are usually not true or false without further qualifications.
   c. terms denoting indefinite degree (e.g., large, long time, regularly), or absolutes (e.g., never, only, always).
   d. placing items in a systematic order (e.g., TTFF, TFTF, and so on).
   e. taking statements directly from the text and presenting them out of context.

MATCHING ITEMS

Like true–false items, matching items represent a popular and convenient testing format. Just like good true–false items, though, good matching items are not as easy to write as you might think. Imagine you are back in your tenth-grade American History class and the following matching item shows up on your test. Is it a good matching exercise or not? If not, what is wrong with it?

Directions: Match A and B.

A	B
1. Lincoln	a. President during the twentieth century
2. Nixon	b. Invented the telephone
3. Whitney	c. Delivered the Emancipation Proclamation
4. Ford	d. Recent president to resign from office
5. Bell	e. Civil rights leader
6. King	f. Invented the cotton gin
7. Washington	g. Our first president
8. Roosevelt	h. Only president elected for more than two terms

See any problems? Compare the problems you have identified with the list of faults and explanations below.

Faults Inherent in Matching Items

Lack of homogeneity   The lists are not homogeneous. Column A contains names of presidents, inventors, and a civil rights leader. Unless specifically taught as a set of related public figures or ideas, this example represents too wide a variety for a

matching exercise. To prevent this from happening you might title your lists (e.g., "U.S. Presidents"). This will help keep irrelevant or filler items from creeping in. If you really want to measure student knowledge of presidents, inventors, and civil rights leaders, then build three separate matching exercises. Doing so will prevent implausible options from being eliminated by the student. When students can eliminate implausible options, they are more likely to guess correctly. For example, the student may not know a president who resigned from office but may know that Washington and Lincoln were presidents, and that neither was recent. Thus the student could eliminate two options, increasing the chance of guessing correctly from one out of eight to one out of six.

Wrong order of lists   The lists should be reversed; that is, Column A should be Column B, and Column B should be Column A. This is a consideration that will save time for the test taker. We are trained to read from left to right. When the longer description is in the left-hand column, the student only reads the description once and glances down the list of names to find the answer. As the exercise is now written, the student reads a name and then has to read through all or many of the more lengthy descriptions to find the answer, a much more time-consuming process.

Easy guessing   There are equal numbers of options and descriptions in each column. Again, this increases the chances of guessing correctly through elimination. In the preceding exercise, if a student did not know who invented the cotton gin but knew which of the names went with the other seven descriptions, the student would arrive at the correct answer through elimination. If there are at least three more options than descriptions, the chances of guessing correctly in such a situation are reduced to one chance in four. Alternatively, the instructions for the exercise may be written to indicate that each option may be used more than once.

Poor directions   Speaking of directions, those included were much too brief. Matching directions should specify the basis for matching. For example,

> DIRECTIONS: Column A contains brief descriptions of historical events. Column B contains the names of presidents. Indicate which man was president when the historical event took place by placing the appropriate letter to the left of the number in Column A.

The original directions also do not indicate how the matches should be shown. Should lines be drawn? Should letters be written next to numbers, or numbers next to letters? Failure to indicate how matches should be marked can greatly increase your scoring time.

Too many correct responses   The description "President during the twentieth century" has three defensible answers: Nixon, Ford, and Roosevelt. You say you meant Henry Ford, inventor of the Model T, not Gerald Ford! Well, that brings us to our final criticism of this matching exercise.

Ambiguous lists   The list of names is ambiguous. Franklin Roosevelt or Teddy Roosevelt? Henry Ford or Gerald Ford? When using names, always include first and last names to avoid such ambiguities.

Now that we have completed our analysis of this test item, we can easily conclude that it needs revision. Let's revise it, starting by breaking the exercise into homogeneous groupings.

DIRECTIONS: Column A describes events associated with U.S. presidents. Indicate which name in Column B matches each event by placing the appropriate letter to the left of the number of Column A. Each name may be used only once.

Column A	*Column B*
_____ 1. A president not elected to office	**a.** Abraham Lincoln
_____ 2. Delivered the Emancipation Proclamation	**b.** Richard Nixon
_____ 3. Only president to resign from office	**c.** Gerald Ford
_____ 4. Only president elected for more than two terms	**d.** George Washington
	e. Franklin Roosevelt
_____ 5. Our first president	**f.** Theodore Roosevelt
	g. Thomas Jefferson
	h. Woodrow Wilson

We can make one more clarification. It is a good idea to introduce some sort of order—chronological, numerical, or alphabetical—to your list of options. This saves the reader time. Students usually go through the list several times in answering a matching exercise, and it is easier to remember a name's or date's location in a list if it is in some sort of order. We can arrange the list of names in alphabetical order to look like this:

Column A	*Column B*
_____ 1. A president not elected to office	**a.** Gerald Ford
_____ 2. Delivered the Emancipation Proclamation	**b.** Thomas Jefferson
_____ 3. Only president to resign from office	**c.** Abraham Lincoln
_____ 4. Only president elected for more than two terms	**d.** Richard Nixon
	e. Franklin Roosevelt
_____ 5. Our first president	**f.** Theodore Roosevelt
	g. George Washington
	h. Woodrow Wilson

Our original exercise contained two items relating to invention. If we were determined to measure only knowledge of inventors through a matching exercise, we would want to add at least one more item. Normally, at least three items are used for matching exercises. Such an exercise might look like the following:

DIRECTIONS: Column A lists famous inventions and Column B famous inventors. Match the inventor with the invention by placing the appropriate letter in the space to the left of the number in Column A. Each name may be used only once.

Column A	*Column B*
_____ **1.** Invented the cotton gin	**a.** Alexander Graham Bell
_____ **2.** One of his inventions was the telephone	**b.** Henry Bessemer
_____ **3.** Famous for inventing the wireless	**c.** Thomas Edison
	d. Guglielmo Marconi
	e. Eli Whitney
	f. Orville Wright

Notice we have complete directions, there are three more options than descriptions, the lists are homogeneous, and the list of names is alphabetically ordered. But what about the final item remaining from our original exercise? Let's say we want to determine whether our students know that Martin Luther King, Jr., was a civil rights leader. We can construct another matching exercise with one column listing the names of civil rights leaders and another listing civil rights accomplishments. However, an alternative would be simply to switch item formats. Usually, single items that are removed from matching exercises because of their lack of homogeneity are easily converted into true–false, completion, or, with a little more difficulty, multiple-choice items. For example,

True–False

T   F   Martin Luther King, Jr., was a civil rights leader.

Completion

The name of the black civil rights leader assassinated in 1968 is _____ .

Multiple Choice

Which one of the following was a civil rights leader?
a. Jefferson Davis
b. Martin Luther King, Jr.
c. John Quincy Adams
d. John Wilkes Booth

Suggestions for Writing Matching Items

1. Keep both the list of descriptions and the list of options fairly short and homogeneous—they should both fit on the same page. Title the lists to ensure homogeneity and arrange the descriptions and options in some logical order. If this is impossible, you're probably including too wide a variety in the exercise. Try two or more exercises.

2. Make sure that all the options are plausible distractors for each description to ensure homogeneity of lists.

3. The list of descriptions should contain the longer phrases or statements and should be on the left, while the options in the right column should consist of short phrases, words, or symbols.

4. Each description in the list should be numbered (each is an item), and the list of options should be identified by letter.

5. Include more options than descriptions. If the option list is longer than the description list, it is harder for students to eliminate options. If the option list is shorter, some options must be used more than once. Always include some options that do not match any of the descriptions, or some that match more than one, or both.

6. In the directions, specify the basis for matching and whether options can be used more than once.

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

_____ 5. Some test items
  a. are too difficult.
  b. are objective.
  c. are poorly constructed.
  d. have multiple defensible answers.
_____ 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
_____ 7. When 53 Americans were held hostage in Iran,
  a. the United States did nothing to try to free them.
  b. the United States declared war on Iran.
  c. the United States first attempted to free them by diplomatic means and later attempted
  d. rescue.
  e. the United States expelled all Iranian students.
_____ 8. The square root of 256 is
  a. 14.
  b. 16.
  c. 4 × 4
  d. both a and c.
  e. both b and c.
  f. all of the above.
  g. 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
  a. is called an objective item.
  b. has content validity.
  c. is too easy.
  d. 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. Hmmm, both a and c? No, that can't be it because I know that 4 × 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.

Higher-level Multiple-Choice Questions

Good multiple-choice items are the most time-consuming kind of objective test items to write. Unfortunately, most multiple-choice items are also written at the knowledge level of the taxonomy of educational objectives. As a new item writer (and, if you're not careful, as an experienced item writer) you will have a tendency to write items at this level. In this section we will provide you with suggestions for writing multiple-choice items to measure higher-level thinking.

The first step is to write at least some objectives that measure comprehension, application, analysis, synthesis, or evaluation to ensure that your items will be at the higher than knowledge level—if your items match your objectives! The following objectives measure behavior at the knowledge level:

The student will name, from memory, the first three presidents of the United States by next Friday.

Given a color chart, the students will identify each of the primary colors.

Objectives such as those will generate multiple-choice items that will measure only memorization. By contrast, the following objectives measure behavior at higher than the knowledge level:

> Given a copy of the President's State of the Union address, the student will be able to identify one example of a simile and one of a metaphor.

> The student will be able to correctly solve three-digit addition problems without regrouping.

With objectives such as those, higher-level multiple-choice items would have to be constructed to match the objectives. Some suggestions for other approaches to measuring at higher than the knowledge level follow.

Use pictorial, graphical, or tabular stimuli Pictures, drawings, graphs, tables, and so on require the student to think at the application level of the taxonomy of educational objectives and may involve even higher levels of cognitive processes. Also, the use of such stimuli can often generate several higher-level multiple-choice items rather than a single higher-level multiple-choice item, as Figure 7.1 illustrates. Other items based on the map in Figure 7.1 could be as follows:

1. Approximately how many miles is it from Dube to Rog?
   a. 100 miles
   b. 150 miles
   c. 200 miles
   d. 250 miles

2. In what direction would someone have to travel to get from Wog to Um?
   a. northwest
   b. northeast
   c. southwest
   d. southeast

A variation on the same theme is to include several pictorial stimuli to represent options and build several stems around them. The following items would be appropriate for a plane geometry class:
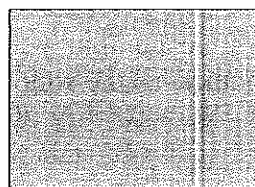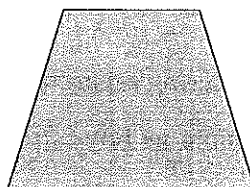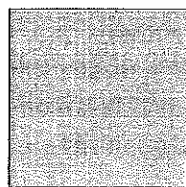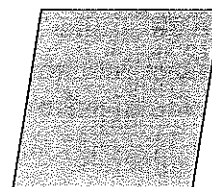

(a)   (b)   (c)

(d)   (e)

In the following questions you are asked to make inferences from the data which are given you on the map of the imaginary country, Serendip. The answers in most instances must be probabilities rather than certainties. The relative size of towns and cities is not shown. To assist you in the location of the places mentioned in the questions, the map is divided into squares lettered vertically from A to E and numbered horizontally from 1 to 5.
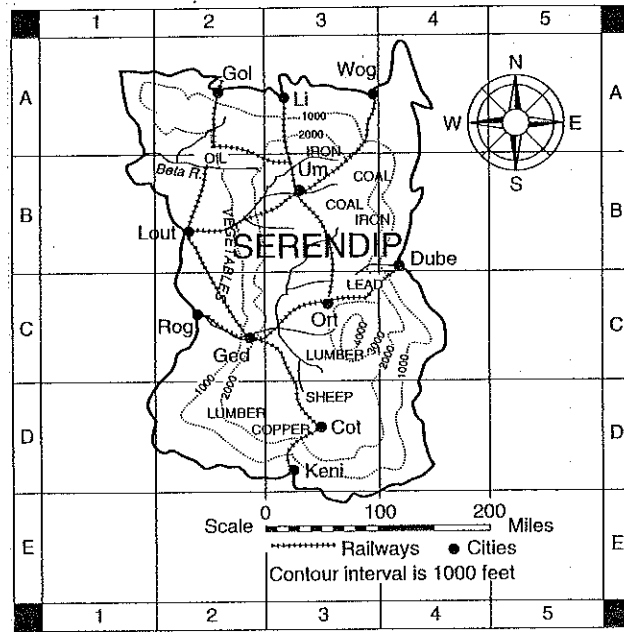
Which of the following cities would be the best location for a steel mill?

(A) Li	(3A)
(B) Um	(3B)
(C) Cot	(3D)
(D) Dube	(4B)

FIGURE 7.1 Use of pictorial stimulus to measure high-level cognitive processes.

1. Which of the figures is a rhombus?
2. Which of the figures is a square?
3. Which of the figures is a trapezoid?

Naturally, it is important to include several more stimulus pictures than items in order to minimize guessing.

Use analogies that demonstrate relationships among terms To answer analogies correctly, students must not only be familiar with the terms, but be able to understand how the terms relate to one another, as the following examples show.

1. Man is to woman as boy is to
   a. father.
   b. mother.
   c. girl.
   d. boy.

2. Physician is to humans as veterinarian is to
   a. fruits.
   b. animals.
   c. minerals.
   d. vegetables.

Require the application of previously learned principles or procedures to novel situations To test whether students really comprehend the implications of a procedure or principle, have the students use the principle or procedure with new information or in a novel way. This requires that the student do more than simply "follow the steps" in solving a problem. It asks the student to demonstrate an ability to go beyond the confines within which a principle or procedure was originally learned.

1. In class we discussed at length Darwin's notion of the "survival of the fittest" within the animal world. Which of the following best describes how this principle applies to the current competitive residential construction industry?
   a. Those builders in existence today are those who have formed alliances with powerful financial institutions.
   b. Only those builders who emphasize matching their homes to the changing structure of the family will survive in the future.
   c. The intense competition for a limited number of qualified home buyers will eventually "weed out" poorly managed construction firms.
   d. Only those home builders who construct the strongest and most durable homes will survive in the long term.

2. (*Follows a lesson on division that relied on computation of grade point averages as examples.*) After filling up his tank with 18 gallons of gasoline, Mr. Watts said to his son, "We've come 450 miles since the last fill-up. What kind of gas mileage are we getting?" Which of the following is the best answer?
   a. 4 miles per gallon.
   b. 25 miles per gallon.
   c. Between 30 and 35 miles per gallon.
   d. It can't be determined from the information given.

These examples are intended to stimulate your creativity—they are by no means exhaustive of the many approaches to measuring higher-level cognitive skills with multiple-choice items. Rather than limit yourself to pictorial items and analogies, use them where appropriate and also develop your own approaches. Remember the main point: Be sure your items match your objectives. Do not write higher-level items if your objectives are at the knowledge level. Doing so will impair your test's content validity. In Chapter 9 we will consider another way you can measure higher-level thinking skills, called performance-based assessment, and we cover yet another, called portfolio assessment, in Chapter 10.

Suggestions for Writing Multiple-Choice Items

Here are some guidelines for writing multiple-choice tests:

1. The stem of the item should clearly formulate a problem. Include as much of the item as possible, keeping the response options as short as possible. However, include only the material needed to make the problem clear and specific. Be concise—don't add extraneous information.

2. Be sure that there is one and only one correct or clearly best answer.

3. Be sure wrong answer choices (distractors) are plausible. Eliminate unintentional grammatical clues, and keep the length and form of all the answer choices equal. Rotate the position of the correct answer from item to item randomly.

4. In most cases it is more important for the student to know what a specific item of information is rather than what it is not. So, only use negative questions or statements if the knowledge being tested requires it, and be sure to italicize or somehow highlight the negative(s).

5. Include from three to five options (two to four distractors plus one correct answer) to optimize testing for knowledge rather than encouraging guessing. It is not necessary to provide additional distractors for an item simply to maintain the same number of distractors for each item. This usually leads to poorly constructed distractors that add nothing to test validity and reliability.

6. To increase the difficulty of a multiple-choice item, increase the similarity of content among the options.

7. Use the option "none of the above" sparingly and only when the keyed answer can be classified unequivocally as right or wrong. Don't use this option when asking for a best answer.

8. Avoid using "all of the above." It is usually the correct answer and makes the item too easy for students with partial information.

Thus far we have considered true–false, matching, and multiple-choice items. We have called these items objective items, but they are also referred to as recognition items. They are recognition items because the test taker needs only to "recognize" the correct answer. Contrast this with "recall" or "supply" formats such as essays and completion items. With essays and completion items, it is much more difficult to guess the right answer than with true–false, matching, or multiple-choice items.

Nevertheless, we will classify completion items with true–false, matching, and multiple-choice items. We call items written in these formats objective items because of the way they are scored, which tends to be fairly straightforward and reliable. This is in contrast to essays, which we will call subjective, because their somewhat less reliable scoring makes them more prone to bias.

COMPLETION ITEMS

Like true–false items, completion items are relatively easy to write. Perhaps the first tests classroom teachers construct and students take are completion tests. Like items of all other formats, though, there are good and poor completion items. Work through the following exercise, again relying on your common sense to identify good and poor items.After having worked through the three previous exercises, you are probably now adept at recognizing common item-writing flaws.

The first item probably reminds you of many you have seen. It is a good rule to avoid using more than one blank per item. The item writer had a specific evolutionary

EXERCISE: *Put a G in the space next to the items you feel are good and a P next to the items you feel are poor.*

_____1. The evolutionary theory of [Darwin] is based on the principle of [survival of the fittest].

_____2. Columbus discovered America in [1492].

_____3. The capital of Mexico is [Mexico City].

_____4. In what year did William J. Clinton become president of the United States? [1992]

_____5. [Too many] blanks cause much frustration in [both test takers and test scorers].

_____6. [Armstrong] was the first American to [walk on the moon].

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

theorist in mind when writing this item, but the final form of the item is not at all focused toward one single theorist. There are a variety of possible correct answers to this item. Not only are such items disturbing and confusing to test takers, they are very time-consuming and frustrating to score. An acceptable revision might look like this:

The evolutionary theory of Darwin is based on the principle of [survival of the fittest].

If you marked item 2 with a G, you were probably in the majority. This is a standard type of completion item that is frequently used. It is also the kind of item that can generate student–teacher conflict. Granted, "1492" is probably the answer most students who studied their lesson would write. But how would you score a response like "a boat" or "the 15th century" or "a search for India"? These may not be the answers you wanted, but they are correct answers.

This illustrates the major disadvantage of completion items and gives you some idea on a much smaller scale of the kinds of difficulties encountered in scoring essay items as well. Unless you take pains to be very specific when you word completion items, you will come across similar situations frequently. In general, it's better to be very specific in writing completion items. Item 2 could be made specific by adding the words "the year," as illustrated here:

Columbus discovered America in the year [1492].

In this form the item leaves little to be interpreted subjectively by the student. The test taker doesn't spend time thinking about what the question is "really" asking, and the test scorer doesn't spend time trying to decide how to score a variety of different, but correct, answers. For once everybody's happy.

Item 3 is a similar case. Consider the following dialogue as heard by one of the authors in a teachers' lounge at an elementary school:

Ms. Rigidity: (*To no one in particular*) Smart-aleck kids nowadays! Ask them a simple question and you get a smart-aleck answer. Kids today don't give you any respect.

Ms. Feelings: I hear some frustration in what you're saying. Anything I can do?

Ms. Rigidity: No, there's nothing you can do, but listen to this. On the last test one of the questions was "The largest city in Puerto Rico is _____." Simple enough, right? Well, not for Mitch! Instead of answering "San Juan," he answered "the capital

city." Smart-aleck kid. We'll see how smart he feels when he gets no credit for that ridiculous answer.

Ms. Feelings: What I hear you saying is that you feel Mitch may not have known that San Juan is the largest city in Puerto Rico.

Ms. Rigidity: Of course he doesn't know—otherwise he would have given the correct answer. That's the whole point!

Author: (*Never known for his tactfulness*) He did give a correct answer. Your question wasn't specific enough.

Ms. Rigidity: I've been using this same test for years, and there are always one or two kids who give me the same answer Mitch did! And they are always kids who lack respect!

Author: How do you know they lack respect?

Ms. Rigidity: Because they always argue with me when they get their tests back. In fact, they say the same thing you did! You're as bad as they are!

Ms. Feelings: I'm hearing some frustration from both of you.

How could this have been avoided? Let's look at Ms. Rigidity's item again.

The largest city in Puerto Rico is _____.

Since San Juan is the largest city in Puerto Rico and is the capital city of Puerto Rico as well, then "the largest city in Puerto Rico is the capital city." There are at least two defensible answers. Just as in true–false, matching, and multiple-choice items, we should strive to avoid multiple defensible answers in completion items. But how can this be avoided? Again, be specific. Made more specific, Ms. Rigidity's item looks like this:

The name of the largest city in Puerto Rico is [San Juan].

Of course, you could have students who carry things to an extreme. For example, some might claim the following is a defensible answer:

The name of the largest city in Puerto Rico is [familiar to many people].

Only you, as the classroom teacher, can determine which answers are defensible and which are not. Your job will be to determine which responses are logical derivations of your test item and which responses are creative attempts to cover up for lack of mastery of content. Keep an open mind, and good luck! Don't be like Ms. Rigidity!

We can clean up item 3 in much the same fashion as we did Ms. Rigidity's item. Revised in this way, item 3 looks like this:

The name of the capital city of Mexico is [Mexico City].

Adding "name of the" to the original item minimizes your chances of students responding that the capital of Mexico is "a city," "very pretty," "huge," "near central Mexico," and so forth.

Item 4 is an example of a well-written completion item. It is specific, and it would be difficult to think of defensible answers other than "1992."

Both the fifth and sixth items illustrate a case of having too many blanks, which prevents the item from taking on any single theme. Blanks are contagious—avoid using more than one.

Suggestions for Writing Completion Items

1. If at all possible, items should require a single-word answer, or a brief and definite statement. Avoid statements that are so indefinite that they may be logically answered by several terms.
   a. Poor item: World War II ended in _____.
   b. Better item: World War II ended in the year _____.

2. Be sure the question or statement poses a problem to the examinee. A direct question is often more desirable than an incomplete statement (it provides more structure).

3. Be sure the answer that the student is required to produce is factually correct. Be sure the language used in the question is precise and accurate in relation to the subject matter area being tested.

4. Omit only key words; don't eliminate so many elements that the sense of the content is impaired.
   a. Poor item: The _____ type of test item is usually more _____ than the _____ type.
   b. Better item: The completion type of test item is usually graded less objectively than the _____ type.

5. Word the statement such that the blank is near the end of the sentence rather than near the beginning. This will prevent awkward sentences.

6. If the problem requires a numerical answer, indicate the units in which it is to be expressed.

You've now used common sense and an increased level of "test-wiseness" to analyze and think about different types of objective test items. In Chapter 8 we will extend our discussion to essay items. Before we move on to essays, however, let's consider one more topic related to item writing, one that applies equally to objective and essay items. This topic is gender and racial bias in test items.

GENDER AND RACIAL BIAS IN TEST ITEMS

An important but often overlooked aspect of item writing involves gender or racial bias. Over the last decades many have become increasingly aware of, and sensitive to, such issues. Professional item writers take great care to eliminate or minimize the extent to which such biases are present in their test items. The classroom teacher would be wise to follow their lead.

One example of gender bias is the exclusive use of the male pronoun *he* in test items. Since the item writer may use it unconsciously, it does not necessarily follow that the item writer is biased. However, this does not prevent the practice from offending a proportion of the population. Similarly, referring exclusively in our items only to members of a single ethnic group will likely be offensive to individuals of different ethnicity. Again, such a practice may be almost unconscious and may not reflect an ethnic bias, but this will not prevent others from taking offense to it.

To avoid such bias, you should carefully balance your references in items. That is, always check your items to be sure that fairly equal numbers of references to males and

females are made. Obviously, equal care and time should be devoted to ensure that ethnic groups are appropriately represented in your items. Such considerations are especially relevant when items are being written at higher than the knowledge level. Since such items often are word problems involving people, gender, and racial bias can easily creep in. The test assembly checklist in Chapter 11 (Figure 11.1) will help remind you to be on the watch for such bias.

Remember, our goal is to measure learning in as valid and reliable a fashion as possible. When emotions are stimulated by gender-biased or racially biased items, these emotions can interfere with valid measurement, leaving us with results that are less useful than they would be otherwise. Given all the care and time we have taken to learn to develop good tests, it makes good sense to take just a bit more to avoid racial and gender bias in our items.

GUIDELINES FOR WRITING TEST ITEMS

In this chapter we have provided you with a good deal of information related to item writing. Much of this information is condensed in the summary, and some general guidelines for item writing are also included. Following these guidelines you will find a summary of the advantages and disadvantages of various item formats. This section should help you make decisions about the type of format to use for your test items. Remember, you will write poor items until the recommendations we have provided become second nature, which comes only with practice. Finally, review the sidebar (Box 7-1) for some suggestions on how a computer can save item-writing time and—with appropriate item-writing software—actually help improve the quality of the items you write.

1. Begin writing items far enough in advance that you will have time to revise them.
2. Match items to intended outcomes at the proper difficulty level to provide a valid measure of instructional objectives. Limit the question to the skill being assessed.
3. Be sure each item deals with an important aspect of the content area and not with trivia.
4. Be sure that the problem posed is clear and unambiguous.
5. Be sure that each item is independent of all other items. The answer to one item should not be required as a condition for answering the next item. A hint to one answer should not be embedded in another item. ⸙
6. Be sure the item has one correct or best answer on which experts would agree.
7. Prevent unintended clues to the answer in the statement or question. Grammatical inconsistencies such as *a* or *an* give clues to the correct answer to those students who are not well prepared for the test.
8. Avoid replication of the textbook in writing test items; don't quote directly from textual materials. You're usually not interested in how well the student memorized the text. Besides, taken out of context, direct quotes from the text are often ambiguous.

BOX

COMPUTERS AND TEST ITEMS

Computers can help save item-writing time through word processing functions and item-writing software. With a word processing program, you can easily make minor or major modifications to test items or even whole tests. You also can reorder the items or change the order of distractors for multiple-choice items much more quickly than you could without a word processing program. This could be a major time saver when it is important to develop alternate forms of a test. Sharing of well-written items is facilitated if they are stored on a flash drive, CD, or hard drive and are made available to all teachers, either by sharing flash drives or other portable media, through a local area network (LAN), or over the Web.

Commercial software is also available that can assist the teacher in constructing test items. By indicating common faults of a test item, these programs minimize the chances that you will construct and use poor items.

Programs are also available that will generate alternate forms of tests according to instructional objectives and difficulty levels. Given a pool of test items and their difficulty levels, the computer could be instructed to make a number of shorter tests based on common objectives and at prespecified difficulty levels, thereby ensuring that different tests would be comparable. With the passage of NCLB and IDEIA, regular classroom teachers will play a greater role in the evaluation of the progress of special learners in the general education curriculum. Because the range of achievement of special learners may vary considerably, and because a variety of accommodations may be needed because of their disabilities, a wider range of item difficulties and formats may be required for special learners than for regular education pupils. The computer may be helpful in this regard, allowing easier and more accurate item storage, modification, and incorporation into tests and assessments of varying lengths and difficulties. Being able to use a computer to efficiently "customize" tests designed for a classroom of regular education students to fit the needs of the special learners will enable you to devote more time to instructional activities. This may minimize the frustration of having to "start from scratch" with each new addition to your class, and may also reduce frustration in special learners and enhance the validity of your assessments of their progress in the general education curriculum.

9. Avoid trick or catch questions in an achievement test. Don't waste time testing how well the students can interpret your intentions.

10. Try to write items that require higher-level thinking.

ADVANTAGES AND DISADVANTAGES OF DIFFERENT OBJECTIVE ITEM FORMATS

True–False Tests
Advantages

Because true–false questions tend to be short, more material can be covered than with any other item format. Thus true–false items tend to be used when a great deal of content has been covered.

True–false questions take less time to construct, but avoid taking statements directly from the text and modifying them slightly to create an item.

Scoring is easier with true–false questions, but avoid having students write "true" or "false" or a "T" or "F." Instead, have them circle "T" or "F" provided for each item.

Disadvantages

True–false questions tend to emphasize rote memorization of knowledge, although sometimes complex questions can be asked using true–false items.

True–false questions presume that the answer to the question or issue is unequivocally true or false. It would be unfair to ask the student to guess at the teacher's criteria for evaluating the truth of a statement.

True–false questions allow for and sometimes encourage a high degree of guessing. Generally, longer examinations are needed to compensate for this.

Matching Tests

Advantages

Matching questions are usually simple to construct and to score.

Matching items are ideally suited to measure associations between facts.

Matching questions can be more efficient than multiple-choice questions because they avoid repetition of options in measuring associations.

Matching questions reduce the effects of guessing.

Disadvantages

Matching questions sometimes tend to ask students trivial information.

They emphasize memorization.

Most commercial answer sheets can accommodate no more than five options, thus limiting the size of any particular matching item.

Multiple-Choice Tests

Advantages

Multiple-choice questions have considerable versatility in measuring objectives from the knowledge to the evaluation level.

Since writing is minimized, a substantial amount of course material can be sampled in a relatively short time.

Scoring is highly objective, requiring only a count of the number of correct responses.

Multiple-choice items can be written so that students must discriminate among options that vary in degree of correctness. This allows students to select the best alternative and avoids the absolute judgments found in true–false tests.

Since there are multiple options, effects of guessing are reduced.

Multiple-choice items are amenable to item analysis (see Chapter 11), which permits a determination of which items are ambiguous or too difficult.

Disadvantages

Multiple-choice questions can be time-consuming to write. If not carefully written, multiple-choice questions can sometimes have more than one defensible correct answer.

Completion Tests

Advantages

Construction of a completion question is relatively easy.

Guessing is eliminated since the question requires recall.

Completion questions take less time to complete than multiple-choice items, so greater amounts of content can be covered.

Disadvantages

Completion questions usually encourage a relatively low level of response complexity.

The responses can be difficult to score since the stem must be general enough so as not to communicate the correct answer. This can unintentionally lead to more than one defensible answer.

The restriction of an answer to a few words tends to measure the recall of specific facts, names, places, and events as opposed to more complex behaviors.

SUMMARY

This chapter introduced you to four types of objective test items: true–false, matching, multiple choice, and completion. The major points are as follows:

1. Choice of item format is sometimes determined by your instructional objectives. At other times the advantages and disadvantages of the different formats should influence your choice.

2. True–false items require less time to construct than other objective items, but they are most prone to guessing, as well as a variety of other faults. These include absolutes in wording, double negatives, opinionated and double-barreled statements, excessive wordiness, and a tendency to reflect statements taken verbatim from readings.

3. Matching items are fairly easy to construct but tend to be subject to the following faults: lack of clarity and specificity in directions, dissimilar and nonordered lists, and reversal of options and descriptions.

4. Multiple-choice items are the most difficult of the objective items to construct. However, higher-order multiple-choice items lend themselves well to measuring higher-level thinking skills. They are subject to several faults, including grammatical cues or specific determiners, multiple defensible answers, unordered option lists, stem clues, opinionated statements, failure to state a problem in the stem, redundant wording, wordiness in the correct option, use of "all of the above," and indiscriminate use of "none of the above."

5. Completion items rival true–false items in ease of construction. Since answers must be supplied, they are least subject to guessing. On the other hand, they require more scoring

time than other objective formats. Common faults include too many blanks, lack of specificity (too many potential responses), and failure to state a problem.

6. To avoid gender and/or racial biases in test items, avoid using stereotypes and be sure to make equal reference to both males and females and to various ethnic groups.

FOR PRACTICE

1. Write a behavioral objective in some area with which you are familiar. Using this objective as your guide, write a test item using each of the four test item formats (true–false, matching, multiple-choice, and completion) discussed in this chapter.

2. Exchange your test items with a classmate. Have him or her check the appropriateness of each of your items for a match with the appropriate objective and against the criteria and guidelines given in this chapter for each test item format. List and correct any deficiencies.

*3. Each of the following items is defective in some way(s). Identify the principal fault or faults in each item and rewrite the item so that it is fault-free.

1. "The Time Machine" is considered to be a
   a. adventure story.
   b. science fiction story.
   c. historical novel.
   d. autobiography.

2. Thaddeus Kosciusko and Casimer Pulaski were heroes in the Revolutionary War. What was their country of origin?
   a. Great Britain
   b. Poland
   c. France
   d. Italy

*Answers to Question 3 appear in Appendix F.

3. The use of force to attain political goals is never justifiable.   (T   F)

4. The personal computer was invented in

   _____ .

5. _____ spent his life trying to demonstrate that

   _____ .

6.
1.	Discovered the Zambezi River	a.	Webb
2.	First female governor	b.	Armstrong
3.	Invented the cotton gin	c.	Minuit
4.	First to swim the English Channel	d.	Livingstone
5.	Purchased Manhattan Island	e.	Whitney
6.	First to walk on the moon	f.	Edison
		g.	Cortez
		h.	Keller
		i.	Rhodes