

## CONSTRUCTING OBJECTIVE TEST ITEMS: SIMPLE FORMS

Each type of test item has its own unique characteristics, uses, advantages, limitations, and rules for construction. In this chapter, these characteristics are considered for objective test forms that typically measure relatively simple learning outcomes: (a) the short-answer item, (b) the true–false item, and (c) the matching exercise.

© CourseSmart

The preliminary test planning described in Chapter 6 provides a sound basis for developing classroom tests that can be used for a variety of instructional purposes. The test specifications clarify the sample of achievement to be measured, and the various considerations in test planning form a general framework within which to proceed. The next step is the actual construction of test items. This step is crucial because the validity of a classroom test is ultimately determined by the extent to which the performance to be measured is actually called forth by the test items. Selecting item types that are inappropriate for the learning outcomes to be measured, constructing items with technical defects, or unwittingly including irrelevant clues in the items can undermine all the careful planning that has gone on before.

The construction of good test items is an art. The skills it requires, however, are the same as those found in effective teaching. Needed are a thorough grasp of subject matter, a clear conception of the desired learning outcomes, a psychological understanding of students, sound judgment, persistence, and a touch of creativity. The only additional requisite for constructing good test items is the skillful application of an array of simple but important rules and suggestions. These techniques of test construction are the topic of this and the next two chapters. The rules for constructing test items, described in these chapters, are applicable to all types of classroom tests using supply-type or selection-type items. Guidelines for constructing assessments using essay questions or other types of performance assessments are provided in Chapters 10 and 11.

In this chapter, we limit our discussion to the simpler forms of objective test items: (a) short-answer items, (b) true-false or alternative-response items, and (c) matching exercises. These item types are treated together, as their use in classroom testing is restricted largely to the measurement of simple learning outcomes in the knowledge area. The discussion of each item type is followed by a checklist for reviewing the items.

## SHORT-ANSWER ITEMS

---

The short-answer item and the completion item both are supply-type test items that can be answered by a word, phrase, number, or symbol. They are essentially the same, differing only in the method of presenting the problem. The short-answer item uses a direct question, whereas the completion item consists of an incomplete statement.

---

### EXAMPLES Short Answer

What is the name of the man who invented the steamboat? (Robert Fulton)

### Completion

The name of the man who invented the steamboat is \_\_\_\_\_. (Robert Fulton)

---

Also included in this category are problems in arithmetic, mathematics, science, and other areas whose solution must be supplied by the student.

## Uses of Short-Answer Items

The short-answer test item is suitable for measuring a wide variety of relatively simple learning outcomes. The following outcomes and test items illustrate some of its common uses.

---

### EXAMPLES Knowledge of Terminology

Lines on a weather map that join points of the same barometric pressure are called \_\_\_\_\_. (isobars)

### Knowledge of Specific Facts

A member of the United States Senate is elected to a term of \_\_\_\_\_ years. (6)

### Knowledge of Principles

If the temperature of a gas is held constant while the pressure applied to it is increased, what will happen to its volume? (It will decrease)

**Knowledge of Method or Procedure**

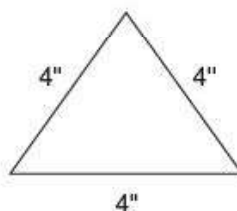
What device is used to detect whether an electric charge is positive or negative? (electroscope)

**Simple Interpretations of Data**

How many syllables are there in the word *Argentina*? (4)

In the number 612, what value does the 6 represent? (600)

In the triangle below, what is the number of degrees in each angle? (60)



If an airplane flying northwest made a 180-degree turn, what direction would it be heading? (southeast)

More complex interpretations can be made when the short-answer item is used to measure the ability to interpret diagrams, charts, graphs, and pictorial data.

Even more notable exceptions to the general rule that short-answer items are limited to measuring simple learning outcomes are found in the areas of mathematics and science, where the solutions to problems can be indicated by numbers or symbols. The following examples illustrate this use.

**EXAMPLES Ability to Solve Numerical Problems**

Milk sells for \$.96 a quart and \$3.68 a gallon. How many cents would you save on each quart of milk if you bought it by the gallon? (4)

**Skill in Manipulating Mathematical Symbols**

If  $\frac{x}{b} = \frac{3}{b-1}$ , then  $x = ?$

$$\frac{3b}{(b-1)}$$

**Ability to Complete and Balance Chemical Equations**

$Mg + (2) HCl \rightarrow ?$

$(MgCl_2 + H_2)$

$(2) Al + (6) HCl \rightarrow ?$

$(2 AlCl_3 + 3H_2)$

For outcomes similar to those in these last examples, the short-answer item is clearly superior to items that require the student to select an answer. The performance described in the learning outcomes is identical with the performance called forth by the



test items. To obtain correct answers, students must actually solve problems, manipulate mathematical symbols, and complete and balance equations.

Attempts are sometimes made to measure such problem-solving activities with selection-type test items, commonly resulting in test items that do not function as intended or that measure quite different learning outcomes. In the following multiple-choice items, for example, note how the division problem can be solved by working it backward (multiplying  $2 \times 43$ , or merely  $2 \times 3$ ) and how in the second problem the value of  $x$  can be determined by substituting each of the alternative answers in the equation on a trial-and-error basis. Such problems obviously do not demand the problem-solving behavior we are attempting to measure.

---

**EXAMPLES** Supply-Type Items Superior to Multiple-Choice (see text):

$$2 \overline{)86} =$$

A 41

B 42

Ⓒ 43

D 44

If  $\frac{x}{4} + \frac{x}{16} = 10$ , then  $x$  equals

A 16

B 24

Ⓒ 32

D 48

---

Similar difficulties are encountered when we substitute selection items measuring the ability to “recognize balanced chemical equations” for short-answer items measuring the ability to “complete and balance chemical equations.” The selection task is a simple one requiring little more than a knowledge of mathematics, but the short-answer task requires extensive knowledge of chemical reactions and their resulting products.

For large-scale testing programs, it is possible to maintain some of the advantages of the supply-type item for mathematics problems that result in numerical answers while still retaining the efficiencies of machine scoring. The grid-in item, for example, was introduced on the Scholastic Assessment Test in the spring of 1994. This item type requires examinees to solve problems and enter their numerical responses in a grid (see example).

---

**EXAMPLE** Five students get scores of 9, 7, 7, 5, and 4 on an eight-item quiz. What is the average score for these five students?

	6	.	4
.	7	7	.
0	0	0	0
1	1	1	1
2	2	2	2
3	3	3	3
4	4	4	4
5	5	5	5
6	6	6	6
7	7	7	7
8	8	8	8
9	9	9	9

Grid-in problems requiring complex problem solving can be constructed, but the approach is likely to be useful only when a test is administered to a large number of students. For classroom use, short-answer problems and extended free-response assessment tasks of the type discussed in Chapters 10 and 11 are likely to provide more information about how students approach such problems and the nature of the difficulties they encounter.

In summary, if the short-answer test item is most effective for measuring a specific learning outcome, then it should be used. We should not discard it for selection-type items unless we are fairly certain that the same learning outcomes will be measured. For many of the simpler learning outcomes, such as knowledge of factual information, changing to some form of selection item will not decrease the validity of the measurement and will result in increased objectivity and ease of scoring. For some of the more complex learning outcomes, such as those in mathematics and science, however, discarding the short-answer test item may mean a change in the learning outcomes being measured and, hence, reduce the validity for the intended outcomes. In deciding whether to use short-answer items or some other item type, our best guide is to follow this principle: Each learning outcome should be measured as directly as possible, and the test-item type most appropriate for the purpose should be used.

## Advantages and Limitations of Short-Answer Items

The short-answer test item is one of the easiest to construct, partly because of the relatively simple learning outcomes it usually measures. Except for the problem-solving outcomes measured in mathematics and science, the short-answer item is used almost exclusively to measure the recall of memorized information.

A more important advantage of the short-answer item is that the students must supply the answer. This reduces the possibility that the students will obtain the correct answer by guessing. They must either recall the information requested or make the necessary computations to solve the problem presented to them. Partial knowledge, which might enable them to choose the correct answer on a selection item, is insufficient for answering a short-answer test item correctly.



Two major limitations restrict the use of the short-answer test item. One—unsuitability for measuring complex learning outcomes—has already been mentioned. The other is the difficulty of scoring. Unless the question is carefully phrased, many answers of varying degrees of correctness must be considered for total or partial credit. For example, a question such as “Where was George Washington born?” could be answered by the name of the city, county, state, region, country, or continent. Although the teacher may have had the name of the state in mind when writing the question, the other answers cannot be dismissed as incorrect. Even when this problem is avoided, however, the scoring may be contaminated by the student’s spelling ability. If full or partial credit is taken off for misspelled words, the students’ test scores will reflect varying degrees of knowledge and spelling skill. If spelling is not counted in the scoring, the teacher must still decide whether misspelled words actually represent the correct answer. We all are familiar with misspellings so bad that it is difficult to determine what the student had in mind. The complications make scoring more time consuming and less objective than that obtained with selection-type items.

These limitations are less troublesome when the answer is to be expressed in numbers or symbols, as in physical science or mathematics. Here, more complex learning outcomes can be measured, spelling is not a problem, and it is usually easier to write test items for which there is only one correct response.

## Suggestions for Constructing Short-Answer Items

The short-answer item is subject to a variety of defects, even though it is considered one of the easiest to construct. The following suggestions will help you avoid possible pitfalls and provide greater assurance that the items will function as intended.

**1.** Word the item so that the required answer is both brief and specific. As indicated earlier, the answer to an item should be a word, phrase, number, or symbol. This can be easily conveyed to the students through the directions at the beginning of the test and by proper phrasing of the question. More difficult is stating the question so that only one answer is correct.

---

### EXAMPLES

*Poor:* An animal that eats the flesh of other animals is (carnivorous).

*Better:* An animal that eats the flesh of other animals is classified as (carnivorous).

---

The first version of this item is so indefinite that it could be completed with answers such as *the wolf*, *the lion*, or even *hungry*. Asking the students to classify this type of animal, as called for in the improved version, better structures the problem and defines the type of response required.

**2.** Do not take statements directly from textbooks to use as a basis for short-answer items. When taken out of context, textbook statements are frequently too general and ambiguous to serve as good short-answer items. Note the vagueness of the first version of the following test item, which was taken verbatim from a chemistry textbook.

**EXAMPLES**

*Poor:* Chlorine is a (halogen).

*Better:* Chlorine belongs to a group of elements that combine with metals to form salts. It is therefore called a (halogen).

Students are most likely to respond to the first version of this test item with the word *gas* because that is the natural state of chlorine, and there is nothing in the statement to imply that the word *halogen* is wanted. The only students who are apt to supply the intended answer are those who memorized the textbook statements. The revised version measures factual knowledge that is not dependent on the phraseology of the textbook. Such items tend to discourage the students from developing little-understood verbal associations based on textbook language and encourage them to achieve the learning outcomes being measured. When items are not taken verbatim from another source, it also forces the item writer to carefully consider the importance of the word or phrase being omitted as well as the other words or phrases in the sentence that might provide a clue to the answer.

3. A direct question is generally more desirable than an incomplete statement. There are two advantages to the direct-question form. First, it is more natural to the students, as this is the usual method of phrasing questions in daily classroom discussions. This is especially important to elementary students when first exposed to short-answer tests. Second, the direct question is usually better structured and free of much of the ambiguity that creeps into items based on incomplete statements. Just the phrasing of a question requires us to decide what it is we want to know.

**EXAMPLES**

*Poor:* John Glenn made his first orbital flight around the earth in \_\_\_\_\_. (1962)

*Better:* When did John Glenn make his first orbital flight around the earth? \_\_\_\_\_. (1962)

*Best:* In what year did John Glenn make his first orbital flight around the earth? \_\_\_\_\_ (1962)

The first version of the item could, of course, be completed with *a space capsule*, *Friendship Seven*, *space*, and similar answers. Putting it in question form forces us to indicate whether it is the time, place, or method we are interested in knowing. The last version is a refinement that makes the question even more specific and that naturally evolves from a consideration of the *when* aspect of the previous question.

4. If the answer is to be expressed in numerical units, indicate the type of answer wanted. For computational problems, it is usually preferable to indicate the units in which the answer is to be expressed. This will clarify the problem and will simplify the scoring.

**EXAMPLES**

*Poor:* If oranges weigh  $5 \frac{2}{3}$  oz. each, how much will a dozen oranges weigh?  
(4 lb. 4 oz.)

*Better:* If oranges weigh  $5 \frac{2}{3}$  oz. each, how much will a dozen oranges weigh? \_\_\_\_\_ lb.  
\_\_\_\_\_ oz. (4) lb. (4) oz.



Unless the type of unit is specified, as in the revised version, correct answers will include 68 oz., 4 1/4 lb., 4.25 lb., and 4 lb. 4 oz. This adds unnecessary confusion to the scoring.

When the problems do not come out even, it is also usually helpful to indicate the degree of precision expected in the answers. For example, specifying that the answers should be “carried out to two decimal places” or “rounded off to the nearest tenth of a percent” makes clear to the students how far to carry their calculations. This will ensure that they reach the degree of precision desired and also prevent them from wasting valuable testing time attempting to achieve a degree of precision not expected.

There are some instances, especially in science, when knowing the proper unit in which the answer is to be expressed and knowing the degree of precision to be expected are important aspects of the learning outcome to be measured. In such cases, the previous suggestions must, of course, be modified.

5. Blanks for answers should be equal in length and in a column to the right of the question. If blanks for answers are kept equal in length, the length of the blank space does not supply a clue to the answer. In the poor version of the following items, the lengths of the blanks restrict the possible answers the students need consider. For the first item they need a long word and for the second item a short one.

---

**EXAMPLES**

*Poor:* What is the name of the part of speech that connects words, clauses, and sentences? \_\_\_\_\_ (conjunction)

What is the name of the part of speech that declares, asserts, or predicts something?  
\_\_\_\_\_ (verb)

*Better:* What is the name of the part of speech that connects words, clauses, and sentences? \_\_\_\_\_ (conjunction)

What is the name of the part of speech that declares, asserts, or predicts something?  
\_\_\_\_\_ (verb)

---

Placing the blanks in a column to the right of the question makes scoring quicker and more accurate. Avoid giving unintentional clues that the answer is a short or long word by making the blanks the same size and large enough for the longest response.

6. When completion items are used, do not include too many blanks. If a statement is too mutilated by blanks, the meaning will be lost, and the student will have to guess what the teacher had in mind. Although some mutilated statements seem to measure complex reasoning abilities, such responses are more appropriate as measures of intelligence than achievement.

---

**EXAMPLES**

*Poor:* (Warm-blooded) animals that are born (alive) and (suckle) their young are called (mammals).

*Better:* Warm-blooded animals that are born alive and suckle their young are called (mammals).

---





## CHECKLIST

## Reviewing Short-Answer Items

	Yes	No
1. Is this the most appropriate type of item to use for the intended learning outcomes?	_____	_____
2. Can the items be answered with a number, symbol, word, or brief phrase?	_____	_____
3. Has textbook language been avoided?	_____	_____
4. Have the items been stated so that only one response is correct?	_____	_____
5. Are the answer blanks equal in length?	_____	_____
6. Are the answer blanks at the end of the items?	_____	_____
7. Are the items free of clues (such as <i>a</i> or <i>an</i> )?	_____	_____
8. Has the degree of precision been indicated for numerical answers?	_____	_____
9. Have the units been indicated when numerical answers are expressed in units?	_____	_____
10. Have the items been phrased so as to minimize spelling errors?	_____	_____
11. If revised, are the items still relevant to the intended learning outcomes?	_____	_____
12. Have the items been set aside for a time before reviewing them?	_____	_____

In the revised version, the blank is at the end of the statement so that the students are presented with a clearly defined problem before they come to the blank. See the “Checklist” box for a review of the guidelines for constructing short-answer tests.

## TRUE-FALSE OR ALTERNATIVE-RESPONSE ITEMS

The alternative-response test item consists of a declarative statement that the student is asked to mark true or false, right or wrong, correct or incorrect, yes or no, fact or opinion, agree or disagree, or the like. In each case, there are only two possible answers. Because the true-false option is the most common, this item type is most frequently referred to as the true-false test item. Some of the variations, however, deviate considerably from the simple true-false pattern and have their own characteristics. For this reason, some prefer the more general category alternative-response item. Here we retain the more commonly used true-false designation.

### Uses of True-False Items

Probably the most common use of the true-false item is in measuring the ability to identify the correctness of statements of fact, definitions of terms, statements of principles, and the like. For measuring such relatively simple learning outcomes, a single declarative statement is used with any one of several methods of responding.

**EXAMPLES** *Directions:* Read each of the following statements. If the statement is true, circle the T. If the statement is false, circle the F.

- |                         |                         |   |
|-------------------------|-------------------------|---|
| <input type="radio"/> T | F                       | 1. The green coloring material in a plant leaf is called chlorophyll. |
| T                       | <input type="radio"/> F | 2. The corolla of a flower includes petals and sepals.                |
| <input type="radio"/> T | F                       | 3. Photosynthesis is the process by which leaves make a plant's food. |

*Directions:* Read each of the following questions. If the answer is yes, circle the Y. If the answer is no, circle the N.

- |                         |                         |   |
|-------------------------|-------------------------|---|
| <input type="radio"/> Y | N                       | 1. Is 50% of 38 more than 18?                             |
| Y                       | <input type="radio"/> N | 2. Is 50% of 4/10 equal to 2/5?                           |
| Y                       | <input type="radio"/> N | 3. If 60% of a number is 9, is the number smaller than 9? |
| <input type="radio"/> Y | N                       | 4. Is 25% of 44 less than 12?                             |

One of the useful functions of the true-false item is in measuring the student's ability to distinguish fact from opinion. The following examples illustrate this use.

**EXAMPLES** *Directions:* Read each of the following statements. If the statement is a fact, circle the F. If the statement is an opinion, circle the O.

- |                         |                         |  |
|-------------------------|-------------------------|--|
| <input type="radio"/> F | O                       | 1. The Constitution of the United States is the highest law of our country.                    |
| F                       | <input type="radio"/> O | 2. The First Amendment to the Constitution is the most important amendment.                    |
| <input type="radio"/> F | O                       | 3. The Fifth Amendment to the Constitution protects people from testifying against themselves. |
| F                       | <input type="radio"/> O | 4. Other countries should adopt a constitution like that of the United States.                 |

*Directions:* Read each of the following statements. If the statement is true, circle the T. If the statement is false, circle the F. If the statement is an opinion, circle the O.

- |                         |                         |                         |   |
|-------------------------|-------------------------|-------------------------|---|
| <input type="radio"/> T | F                       | O                       | 1. The earth is a planet.   |
| T                       | <input type="radio"/> F | O                       | 2. The earth revolves around the moon.                                      |
| T                       | F                       | <input type="radio"/> O | 3. There are intelligent life forms on planets orbiting some distant stars. |

These items measure a learning outcome important to all subject-matter areas: If people are to think critically about a topic, they must first be able to distinguish fact from opinion.

All too frequently, true-false tests include numerous opinion statements to which the student is asked to respond true or false. This is extremely frustrating because there is no objective basis for determining whether a statement of opinion is true or false. The student must usually guess what opinion the teacher holds and mark the answers accordingly. This, of course, is undesirable from all standpoints—testing, teaching, and learning. It is much better to have the student identify statements of opinion as such. An alternative



procedure is to attribute the opinion to some source, making it possible to mark the statements true or false and measuring knowledge concerning the beliefs held by an individual or the values supported by an organization or institution.

© CourseSmart

**EXAMPLES** *Directions:* Read each of the following statements. If the statement is true, circle the T. If the statement is false, circle the F.

- |                                    |                         |  |
|------------------------------------|-------------------------|--|
| T                                  | <input type="radio"/> F | 1. Franklin D. Roosevelt believed that labor unions interfered with the U.S. free-enterprise system.                       |
| T                                  | <input type="radio"/> F | 2. The National Rifle Association favors strict gun control laws.  |
| <input checked="" type="radio"/> T | F                       | 3. The National Education Association opposes the use of public funds for vouchers for students to attend private schools. |

Items like these can become measures of understanding if the opinion statements attributed to an individual or group are new to the student. The task then becomes one of interpreting the beliefs held by the individual or group and applying them to the new situation.

Another aspect of understanding that can be measured by the true-false item is the ability to recognize cause-and-effect relationships. This type of item usually contains two true propositions in one statement, and the student is to judge whether the relationship between them is true or false.

**EXAMPLES** *Directions:* In each of the following statements, both parts of the statement are true. You are to decide whether the second part explains why the first part is true. If it does, circle Yes. If it does not, circle No.

- |                                      |                          |   |
|--------------------------------------|--------------------------|---|
| Yes                                  | <input type="radio"/> No | 1. Leaves are essential <i>because</i> they shade the tree trunk.                         |
| Yes                                  | <input type="radio"/> No | 2. Whales are mammals <i>because</i> they are large.                                      |
| <input checked="" type="radio"/> Yes | No                       | 3. Some plants do not need sunlight <i>because</i> they get their food from other plants. |

© CourseSmart

The true-false item also can be used to measure some simple aspects of logic, as illustrated by the following items that were developed for use in a science test.

**EXAMPLES** *Directions:* Read each of the following statements. If the statement is true, circle the T; if it is false, circle the F. Also, if the converse of the statement is true, circle the CT; if the converse is false, circle the CF. Be sure to give two answers for each statement.

- |                                    |                         |                                     |                                     |  |
|------------------------------------|-------------------------|-------------------------------------|-------------------------------------|--|
| <input checked="" type="radio"/> T | F                       | CT                                  | <input checked="" type="radio"/> CF | 1. All trees are plants.                 |
| T                                  | <input type="radio"/> F | CT                                  | <input checked="" type="radio"/> CF | 2. All parasites are animals.            |
| T                                  | <input type="radio"/> F | <input checked="" type="radio"/> CT | CF                                  | 3. All eight-legged animals are spiders. |
| <input checked="" type="radio"/> T | F                       | <input checked="" type="radio"/> CT | CF                                  | 4. No spiders are insects.               |

A common criticism of the true–false item is that a student may be able to recognize a false statement as incorrect but still not know what is correct. For example, when students answer the following item as false, it does not indicate whether they know what negatively charged particles of electricity are called; all the answer tells us is that they know they are not called neutrons.

---

**EXAMPLE**    T    ☒ F    Negatively charged particles of electricity are called neutrons.

---

This is a rather crude measure of knowledge because there is an infinite number of things that negatively charged particles of electricity are *not* called. To overcome such difficulties, some teachers prefer to have the students change all false statements to true. When this is required, the part of the statement it is permissible to change should be indicated.

---

**EXAMPLES**    *Directions:* Read each of the following statements. If a statement is true, circle the T. If a statement is false, circle the F and change the underlined word to make the statement true. Place the new word in the blank space after the F.

- |                                    |                                    |             |  |
|------------------------------------|------------------------------------|-------------|--|
| T                                  | <input checked="" type="radio"/> F | (electrons) | 1. Particles of negatively charged electricity are called <i>neutrons</i> .              |
| <input checked="" type="radio"/> T | F                                  | _____       | 2. Mechanical energy is turned into electrical energy by means of the <i>generator</i> . |
| T                                  | <input checked="" type="radio"/> F | (store)     | 3. An electric condenser is used to <i>generate</i> electricity.                         |
- 

Unless the key words to be changed are indicated, students are liable to rewrite the entire statement. In addition to the increase in scoring difficulty, this frequently leads to true statements that deviate considerably from the original intent of the item. A clever student may even change false statements to true by simply adding *not* in the appropriate place.

## Advantages and Limitations of True–False Items

A major advantage of true–false items is that they are efficient. Students can typically respond to roughly three true–false items in the time it takes to respond to two multiple-choice items (Ebel & Frisbie, 1991).

Proponents of true–false items such as Ebel and Frisbie (1991) argue that verbal knowledge is central to educational achievement and that “all verbal knowledge can be expressed in propositions” which can be judged to be true or false (p. 135). They make a strong case that true–false items have utility for measuring a broad range of verbal knowledge.

One advantage cited frequently for true–false items is, unfortunately, more illusory than real: ease of construction. This has probably resulted from the common practice of taking statements from textbooks, changing half of them to false statements, and submitting the product to students as a true–false test. Such test items are often so obvious that everyone gets them correct or so ambiguous that even the better students are confused by them. In short, it is easy to construct *poor* true–false items. To construct unambiguous true–false items that measure significant learning outcomes, however, requires much skill.



A second advantage attributed to the true-false item is that a wide sampling of course material can be obtained. Certainly a student can respond to many test items in a short time, which makes it possible to cover a wide range of content, but some types of subject matter do not lend themselves to true-false types of items. True-false statements require course material that can be phrased so that the statements are true or false without qualification or exception. There are areas in which such absolutely true or false statements cannot be made. In some fields, such as the social sciences, practically all significant statements require some qualification. In some subject areas, only relatively trivial statements can be reduced to absolute terms.

One of the most serious limitations of the true-false item is in the types of learning outcomes that can be measured. True-false items are not especially useful beyond the knowledge area. The exceptions to this seem to be distinguishing between fact and opinion and identifying cause-and-effect relationships. These two outcomes are probably the most important measured by this type of item. Many of the learning outcomes measured by the true-false item can be measured more effectively by other forms of selection items, especially the multiple-choice form.

Another factor that limits the usefulness of the true-false item is its susceptibility to guessing. With only two alternatives, a student has a 50/50 chance of selecting the correct answer on the basis of chance alone, and because of the difficulty of constructing items that do not contain clues to the answer, the student's chances of guessing correctly are usually much greater than 50%. This disadvantage is offset, however, by the relatively large number of items that can be answered in a given period of time. Nevertheless, with a typical 100-item true-false test, it is not unusual to have the lowest score above 80. An indeterminate amount of knowledge is reflected in such a score: Many of the correct answers can be accounted for by chance or the presence of clues. A scoring formula utilizing a correction for guessing is frequently suggested as a solution for this problem. This formula compensates only chance guesses, however, and does not include those guided by clues. In addition, such a scoring formula favors individuals willing to take a chance. Even when warned that there will be a penalty for guessing, these individuals will continue to guess, using any clues available, and will do better than chance. Cautious students, on the other hand, will mark only those answers they are certain are correct and will omit many items they could have marked correctly using clues and partial information. Thus, the scores tend to reflect personality differences as well as knowledge of the subject.

The great likelihood of successful guessing on the true-false item has two implications that should be taken into account: (a) The reliability of each item is low, making it necessary to include many items in order to obtain a reliable measure of achievement; and (b) the diagnostic value of such a test is practically nil because analyzing a student's response to each item is meaningless.

One last caution that needs to be considered in the design of tests with true-false items is student response sets. As noted earlier, a response set is a consistent tendency to follow a certain pattern in responding to test items. In taking a true-false test, for example, some students will consistently mark "true" those items they do not know, and others will consistently mark them "false." Thus, if there is not a balance between true and false items, a given test will favor one response set over another and introduce an element into the test score that is irrelevant to the purpose of the test.

True-false items are most useful in situations in which there are only two possible alternatives (e.g., right, left; more, less; who, whom) and special uses such as distinguishing fact



from opinion, cause from effect, superstition from scientific belief, relevant from irrelevant information, valid from invalid conclusions, and the like.

## Suggestions for Constructing True–False Items

The main task in constructing true–false items is formulating statements free from ambiguity and irrelevant clues. This is extremely difficult, and the only guidance that can be given is a list of things to **avoid** when phrasing the statements.

1. Avoid broad general statements if they are to be judged true or false. Most broad generalizations are false unless qualified, and the use of qualifiers provides clues to the answer.

### EXAMPLES

Poor: T (F)  
 Poor: (T) F

The president of the United States is elected to that office.

The president of the United States is usually elected to that office.

In this example, the first version is generally true but must be marked false because there are exceptions, such as when the vice president takes office in event of the president's death. In the second version, the qualifier **usually** makes the statement true but provides a definite clue. Words such as **usually**, **generally**, **often**, and **sometimes** are more likely to appear in true statements, and absolute terms such as **always**, **never**, **all**, **none**, and **only** are more apt to appear in false statements. Although the influence of such clues sometimes can be offset by balancing their use in true–false statements, the simplest solution seems to be to avoid the use of broad generalizations that are obviously false or must be qualified by specific determiners.

2. Avoid trivial statements. In an attempt to obtain statements that are unequivocally true or false, we sometimes inadvertently turn to specific statements of fact that fit this criterion beautifully but have little significance from a learning standpoint.

### EXAMPLES

Poor: (T) F  
 Poor: T (F)

Harry S. Truman was the thirty-third president of the United States.

The United States declared war on Japan on December 7, 1941.

The first item calls for a relatively unimportant fact concerning Truman's tenure as president, and the second item expects the student to remember that the United States did not declare war until December 8. Such items cause students to direct their attention toward memorizing minutiae at the expense of more general knowledge and understanding.

3. Avoid the use of negative statements, especially double negatives. Students tend to overlook negative words such as **no** or **not**, and double negatives contribute to the statement's ambiguity. Note the ambiguity in this relatively simple statement, which uses two negatives.

### EXAMPLES

Poor: (T) F  
 Better: (T) F

None of the steps in the experiment was unnecessary.

All of the steps in the experiment were necessary.



When a negative word must be used, it should be underlined or put in italics so that students do not overlook it.

**4. Avoid long, complex sentences.** As noted earlier, a test item should indicate whether a student has achieved the knowledge or understanding being measured. Long, complex sentences tend also to measure the extraneous factor of reading comprehension and therefore should be avoided in tests designed to measure achievement.

<b>EXAMPLES</b>	<i>Poor:</i> <input checked="" type="radio"/> T <input type="radio"/> F	Despite the theoretical and experimental difficulties of determining the exact pH value of a solution, it is possible to determine whether a solution is acid by the red color formed on litmus paper when it is inserted into the solution.
	<i>Better:</i> <input checked="" type="radio"/> T <input type="radio"/> F	Litmus paper turns red in an acid solution.

As in the preceding example, it frequently is possible to shorten and simplify a statement by eliminating nonfunctional material and restating the main idea. If this is not possible, it may be necessary to change to another item form in order to avoid a complex sentence structure.

**5. Avoid including two ideas in one statement, unless cause-and-effect relationships are being measured.** Some difficulties arising from the inclusion of two ideas in one statement are apparent in the following example, which is one of many similar items a teacher actually used in a biology examination. In each instance, the students were asked to judge merely whether the statement was true or false.

<b>EXAMPLE</b>	<i>Poor:</i> <input type="radio"/> T <input checked="" type="radio"/> F	A worm cannot see because it has simple eyes.
----------------	---	---

This item is keyed false because a worm does not have simple eyes. However, when this teacher asked a student why he marked it false, the student said, "Worms *can* see." This demonstrates that students can get items correct with erroneous information. This is so because the first proposition can be true or false, the second proposition can be true or false, and the relationship between them can be true or false. Thus, when students mark the item false, there is no way of determining to which of the three elements they are responding. The best solution to this dilemma seems to be to use only true propositions and to ask the students to judge the truth or falsity of the relationships between them. Such items also might, of course, be divided into two simple statements, each containing a single idea.

**6. If opinion is used, attribute it to some source, unless the ability to identify opinion is being specifically measured.** Statements of opinion cannot be marked true or false, and it is unfair to expect students to guess how the teacher will score such items or to respond to opinion statements as statements of fact. Knowing whether some significant individual or group supports or refutes a certain opinion, however, can be important from a learning standpoint.

## EXAMPLES

*Poor:* T (F)

Adequate medical care can be best provided through socialized medicine.

*Better:* T (F)

The American Medical Association favors socialized medicine as the best means of providing adequate medical care.

The first version cannot be answered true or false. It may serve a useful purpose in an attitude test, but there is no factual basis on which to decide the truth or falsity of the statement. The second version is clearly false.

7. True statements and false statements should be approximately equal in length. There is a natural tendency for true statements to be longer because such statements must be precisely phrased in order to be absolutely true. This can be overcome by lengthening the false statements through the use of qualifying phrases similar to those found in true statements. Thus, the length of the statement will be eliminated as a possible clue to the correct answer.

8. The number of true statements and false statements should be approximately equal. Constructing a test with an approximately equal number of true statements and false statements will prevent response sets from unduly inflating or deflating the students' scores. You will recall that some students consistently mark statements "true" when in doubt about an answer, whereas others consistently mark them "false." Neither response set should be favored by overloading the test with items of one type.

In honoring this suggestion, the words *approximately equal* should be given special attention. If a teacher consistently uses exactly the same number, this will provide a clue to the student who is unable to answer some of the test items. The best procedure seems to be to vary the percentage of true statements somewhere between 40% and 60%. Under no circumstances should the statements be all true or all false. Students who detect this as a possibility can obtain perfect scores on the basis of one guess.

See the "Checklist" box for questions to use in reviewing true-false items.

## MATCHING EXERCISES

In its traditional form, the matching exercise consists of two parallel columns with each word, number, or symbol in one column being matched to a word, sentence, or phrase in the other column. The items in the column for which a match is sought are called premises, and the items in the column from which the selection is made are called responses. The basis for matching premises to responses is sometimes self-evident but more often must be explained in the directions. In any event, the student's task is to identify the pairs of items that are to be associated on the basis indicated. For example, the student may be asked to identify important historical events, as in the following illustration.





## CHECKLIST

## Reviewing True-False Items

	Yes	No
1. Is this the most appropriate type of item to use?	_____	_____
2. Can each statement be clearly judged true or false?	_____	_____
3. Have specific determiners (e.g., <i>usually</i> , <i>always</i> ) been avoided?	_____	_____
4. Have trivial statements been avoided?	_____	_____
5. Have negative statements (especially double negatives) been avoided?	_____	_____
6. Have the items been stated in simple, clear language?	_____	_____
7. Are opinion statements attributed to some source?	_____	_____
8. Are the true and false items approximately equal in length?	_____	_____
9. Is there an approximately equal number of true and false items?	_____	_____
10. Has a detectable pattern of answers (e.g., T, F, T, F) been avoided?	_____	_____
11. If revised, are the items still relevant to the intended learning outcomes?	_____	_____
12. Have the items been set aside for a time before reviewing them?	_____	_____

© CourseSmart

**EXAMPLE** *Directions:* On the line to the left of each United States space event in Column A, write the letter of the astronaut in Column B who achieved that honor. Each name in Column B may be used once, more than once, or not at all.

Column A	Column B
(G) 1. First United States astronaut to ride in a space capsule	A. Edwin Aldrin
(E) 2. First United States astronaut to orbit the earth	B. Neil Armstrong
(H) 3. First United States astronaut to walk in space	C. Frank Borman
	D. Scott Carpenter
	E. John Glenn
	F. Wally Schirra
(B) 4. First United States astronaut to step on the moon	G. Alan Shepard
	H. Edward White

This matching exercise illustrates an imperfect match; that is, there are more names in column B than are needed to match each event in column A. The directions also indicate that an item may be used once, more than once, or not at all. Both of these procedures prevent students from matching the final pair of items on the basis of elimination.

Two other factors are notable in our example. First, the items in the list of premises in Column A are homogeneous, as they all are concerned with important space events. Such homogeneity is necessary if a matching exercise is to function properly. Second, for each premise in Column A, there are several plausible responses in Column B. Thus, the incorrect responses serve as attractive choices for those students who are in doubt about the correct answers. Both factors tend to minimize the opportunity for successful guessing.

## Uses of Matching Exercises

The typical matching exercise is limited to measuring factual information based on simple associations. Whenever learning outcomes emphasize the ability to identify the relationship between two things and a sufficient number of homogeneous premises and responses can be obtained, a matching exercise seems most appropriate. It is a compact and efficient method of measuring such simple knowledge outcomes. Examples of relationships considered important by teachers, in a variety of fields, include the following:

Persons . . . . .	Achievements
Dates . . . . .	Historical Events
Terms . . . . .	Definitions
Rules . . . . .	Examples
Symbols . . . . .	Concepts
Authors . . . . .	Titles of Books
Foreign Words . . . . .	English Equivalents
Machines . . . . .	Uses
Plants or Animals . . . . .	Classification
Principles . . . . .	Illustrations
Objects . . . . .	Names of Objects
Parts . . . . .	Functions

The matching exercise has also been used with pictorial materials in relating pictures and words or to identify positions on maps, charts, and diagrams. Regardless of the form of presentation, the student's task is essentially to relate two things that have some logical basis for association. This restricts the use of the matching exercise to a relatively small area of student achievement.

## Advantages and Limitations of Matching Exercises

The major advantage of the matching exercise is its compact form, which makes it possible to measure a large amount of related factual material in a relatively short time. This is a mixed blessing, however, as it frequently leads to the excessive use of matching exercises and a corresponding overemphasis on the memorization of simple relationships.

Another advantage often cited for the matching exercise is ease of construction. Poor-matching items can be rapidly constructed, but good-matching items require a high degree of skill. The correct response for each premise must also serve as a plausible response for the other premises. Any lack of plausibility will reduce the number of possible choices and



provide clues to the correct answer. The matching exercise tends to have more irrelevant clues than any other item type, with the possible exception of the true-false item.

The main limitations of the matching exercise are that it is restricted to the measurement of factual information based on rote learning and that it is highly susceptible to the presence of irrelevant clues. Another limitation, somewhat related, is the difficulty of finding homogeneous material that is significant from the viewpoint of our objectives and learning outcomes. For example, we might start out with a few great scientists and their achievements, which we feel all students should know. In order to construct a matching item, it becomes necessary to add the names and achievements of other, lesser-known scientists. Thus, we find ourselves measuring factual information that was not included in our original test plan and that is far less important than other aspects of knowledge we had intended to include. In short, less significant material is introduced into the test because enough significant, homogeneous material is unavailable. This is a common problem in constructing matching exercises and one not easily avoided. One solution is to begin with multiple-choice items, because each item can be directly related to a particular outcome, and to switch to the matching form only when homogeneous material makes the matching exercise a more efficient method of measuring the same achievement.

### Suggestions for Constructing Matching Exercises

Although the matching exercise has only limited usefulness in classroom tests, whenever it is used, special efforts should be made to remove irrelevant clues and to arrange it so that the student can respond quickly and without confusion. The following suggestions are designed to guide such efforts.

1. Use only homogeneous material in a single matching exercise. This has been mentioned before and is repeated here for emphasis. It is without a doubt the most important rule of construction and yet the one most commonly violated. One reason for this is that homogeneity is a matter of degree, and what is homogeneous to one group may be heterogeneous to another. For example, let us assume that we are following the usual suggestion for obtaining homogeneity and develop a matching exercise that includes only men and their achievements. We might end up with a test exercise such as the following one.

**EXAMPLE** *Directions:* On the line to the left of each achievement listed in Column A, write the letter of the man's name in Column B who is noted for that achievement. Each name in Column B may be used once, more than once, or not at all.

Column A	Column B
(A) 1. Invented the telephone	A. Alexander Graham Bell
(B) 2. Discovered America	B. Christopher Columbus
(C) 3. First United States astronaut to orbit the earth	C. John Glenn
(E) 4. First president of the United States	D. Abraham Lincoln
	E. Ferdinand Magellan
	F. George Washington
	G. Eli Whitney



Although the matching exercise in our example may be homogeneous for most students in the primary grades, the discriminations called for are so gross that students above that level will see it as a heterogeneous collection of inventors, explorers, and presidents. Thus, to obtain homogeneity at higher grade levels, it is necessary to have only inventors and their inventions in one matching exercise, explorers and their discoveries in another, and presidents and their achievements in another. At a still higher level, it may be necessary to limit matching exercises still further, such as to inventors whose inventions are in the same field, in order to keep the material homogeneous and free from irrelevant clues. As we increase the level of discrimination called for in a matching exercise, significant homogeneous material becomes increasingly difficult to obtain. Take inventors, for example. How many significant inventions are there in any one area?

**2.** Include an unequal number of responses and premises and instruct the student that responses may be used once, more than once, or not at all. This will make all the responses eligible for selection for each premise and will decrease the likelihood of successful guessing. When an equal number of responses and premises are used and each response is used only once, the probability for guessing the remaining responses correctly is increased each time a correct answer is selected. The odds for correct guessing increase as the list of available responses decreases, and the final response, of course, can be selected entirely on the basis of this process of elimination. In most matching exercises, imperfect matching can be obtained by including more or fewer responses than premises. In either case, the directions should instruct the student that each response may be used once, more than once, or not at all.

**3.** Keep the list of items to be matched brief and place the shorter responses on the right. A brief list of items is advantageous to both the teacher and the student. From the teacher's standpoint, it is easier to maintain homogeneity in a brief list. In addition, there is a greater likelihood that the various learning outcomes and subject-matter topics will be measured in a balanced manner. Because each matching exercise must be based on homogeneous material, a long list will require excessive concentration in one area. From the students' viewpoint, a brief list enables them to read the responses rapidly and without confusion. Approximately four to seven items in each column seems best. There certainly should be no more than 10 items in either column.

Placing the shorter responses on the right also contributes to more efficient test taking, as it enables students to read the longer premise first and then to scan rapidly the list of responses.

**4.** Arrange the list of responses in logical order, place words in alphabetical order, and numbers in sequence. This will contribute to the ease with which the students can scan the responses in searching for the correct answers. It will also prevent them from detecting possible clues from the arrangement of the responses.

---

**EXAMPLE** *Directions:* On the line to the left of each historical event in Column A, write the letter from Column B that identifies the time period when the event occurred. Each date in Column B may be used once, more than once, or not at all.

*Column A*

(B) 1. Boston Tea Party  
(A) 2. Repeal of the Stamp Act

*Column B*

A. 1765–1769  
B. 1770–1774



- |   |              |
|---|--------------|
| (E) 3. Enactment of the Northwest Ordinance | C. 1775–1779 |
| (C) 4. Battle of Lexington                  | D. 1780–1784 |
| (A) 5. Enactment of Townshend Acts          | E. 1785–1789 |
| (B) 6. First Continental Congress           |              |
| (E) 7. United States Constitution drawn up  |              |

This matching exercise also demonstrates the use of fewer responses than premises and the desirability of placing the shortest items on the right.

5. Indicate in the directions the basis for matching the responses and premises. Although the basis for matching is rather obvious in most matching exercises, there are advantages in clearly stating it. First, ambiguity and confusion will be avoided. Second, testing time will be saved because the student will not need to read through the entire list of premises and responses and then “reason out” the basis for matching.

Special care must be taken when stating directions for matching items. Directions that precisely indicate the basis for matching frequently become long and involved, placing a premium on reading comprehension. For younger students, it may be desirable to give oral directions, put an example on the blackboard, and have the students draw lines between the matched items rather than transfer letters.

6. Place all the items for one matching exercise on the same page. This will prevent the disturbance created by 30 or so students switching the pages of the test back and forth. It also will prevent them from missing the responses appearing on another page and generally adds to the speed and efficiency of test administration.

See the “Checklist” box for a review of all the construction guidelines given here.



## CHECKLIST

### Reviewing Matching Items

	Yes	No
1. Is this the most appropriate type of item to use?	___	___
2. Is the material in the two lists homogeneous?	___	___
3. Is the list of responses longer or shorter than the list of premises?	___	___
4. Are the responses brief and on the right-hand side?	___	___
5. Have the responses been placed in alphabetical or numerical order?	___	___
6. Do the directions indicate the basis for matching?	___	___
7. Do the directions indicate that each response may be used more than once?	___	___
8. Is all of each matching item on the same page?	___	___
9. If revised, are the items still relevant to the intended learning outcomes?	___	___
10. Have the items been set aside for a time before reviewing them?	___	___



## SUMMARY

---

The construction of classroom tests, like other phases of teaching, is an art that must be learned. It is not enough to have a knowledge of subject matter, defined learning outcomes, or a psychological understanding of the students' mental processes, although all these are prerequisites. The ability to construct high-quality test items requires a knowledge of the principles and techniques of test construction and skill in their application.

In this chapter, we discussed techniques for constructing short-answer items, true-false or alternative-response items, and matching exercises. Although these simple forms of objective test items can be made suitable for measuring understanding, thinking skills, and other complex achievements, considerable skill in item construction is required to go beyond simple knowledge outcomes. Other types of items and tasks are generally more satisfactory for measuring higher-level skills. Thus, the primary use of these simple types of items is in measuring knowledge outcomes.

The short-answer item requires students to supply the appropriate word, phrase, number, or symbol to a direct question or incomplete statement. It can be used for measuring a variety of simple knowledge outcomes, but it is especially useful for measuring problem-solving ability in science and mathematics. The ease with which short-answer items can be constructed and their relative freedom from guessing favor their use. However, the areas in which they can be effectively used are restricted by the relatively simple learning outcomes measured and by the fact that the scoring can be contaminated by spelling errors. When short-answer items are used, the question must be stated clearly and concisely, be free from irrelevant clues, and require an answer that is both brief and definite. Problems requiring only a number or a symbol for an answer are particularly adaptable to the short-answer form.

The true-false item requires the student to select one of two possible answers. This item type is used for measuring knowledge outcomes when only two alternatives are possible or the ability to identify the correctness of statements of fact is important. It is also adaptable to measuring the ability to distinguish fact from opinion and the ability to recognize cause-and-effect relationships. The use of true-false items is limited by the difficulty of constructing clue-free items that measure significant learning outcomes, the susceptibility of this type to guessing, the low reliability of each item, and the general lack of diagnostic value. When the true-false item is used, special efforts must be made to formulate statements that are free from ambiguity, specific determiners, and clues.

The matching exercises consist of two parallel columns of phrases, words, numbers, or symbols that must be matched. Examples of items included in matching exercises are persons and achievements, dates and historical events, and terms and definitions. The nature of the matching exercise limits it to measuring the ability to identify the relationship between two things. For this restricted use, it is a compact item type that can be used to measure many relationships in a short time. Its limitations include the difficulty of removing irrelevant clues and the difficulty of finding significant homogeneous material. When homogeneous material is available, including more items in one column than in the other, arranging the shorter responses on the right and in logical order, and indicating clearly the basis for matching all will contribute to the effectiveness of the matching exercise.