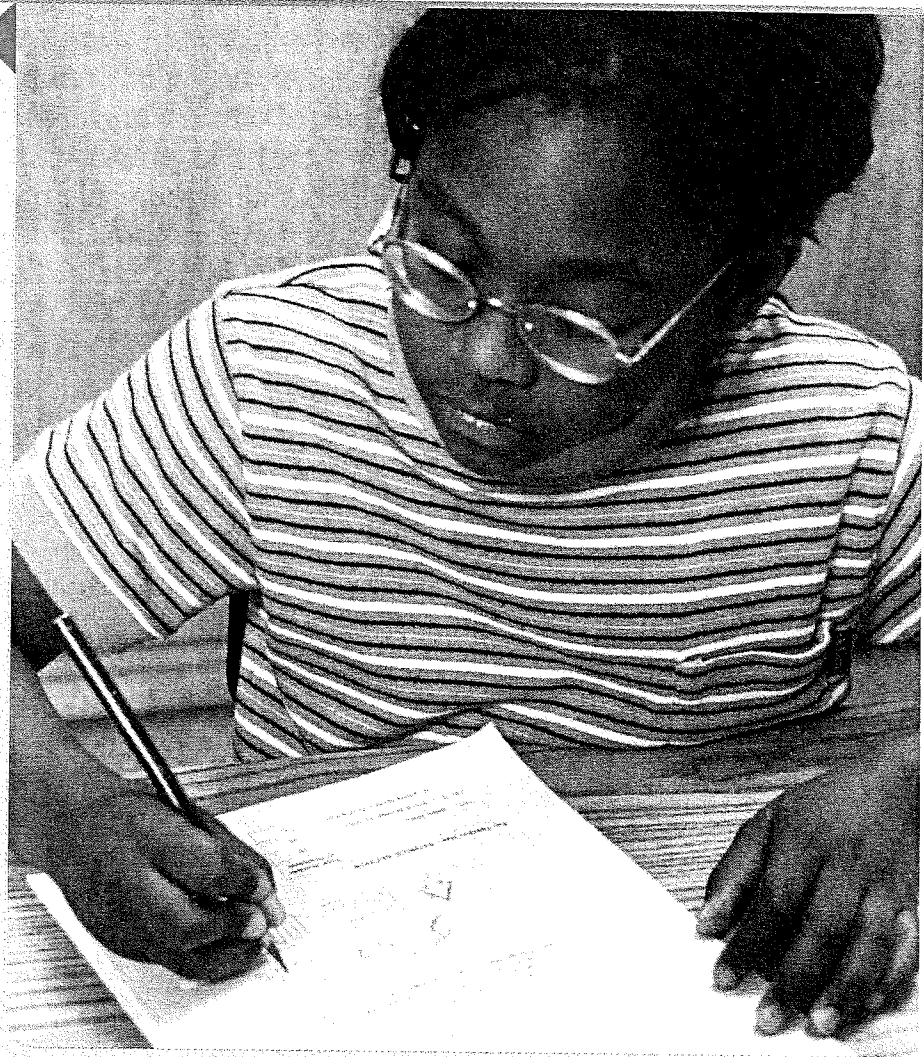


# PREPARING SELECTED AND CONSTRUCTED RESPONSE TESTS

## CHAPTER OBJECTIVES

After reading this chapter, you will be able to:

- Define summative assessment, selected response test items, constructed response test items, item stem, distractors, and specific determiners
- Compare the purpose for and other aspects of diagnostic, formative, and summative assessments
- Write fair and accurate selected response and constructed response test items
- Compare the advantages and disadvantages of different types of selected response and constructed response test items
- List guidelines for constructing or selecting interpretive exercises
- Cite guidelines for writing and critiquing test items
- Describe decisions required in planning a test
- State guidelines for assembling a test



In preparation for her first year of teaching Grade 5, Rita is planning a science unit on weather. As part of this preparation, Rita is developing an assessment plan for this unit that will include different ways of assessing her students' knowledge and understanding of concepts related to weather. Rita has included an end-of-unit test as one of the methods of assessment to be included in her plan. She is committed to constructing a test that will measure her students' higher-level thinking skills as well as their knowledge of concepts related to weather. In order to accomplish this, Rita plans to make use of a variety of different types of test items including multiple-choice, true-false, matching, short-answer, completion, essay, and interpretive exercise questions. She is hopeful that this test will be both interesting to her students and a valid measure of their knowledge about weather and their ability to employ higher-level thinking skills.

At most grade levels, tests are the most commonly used procedure for gathering evidence about student learning. These tests may be constructed by teachers, textbook publishers, Ministries or Departments of Education, or standardized test publishers. In terms of student evaluation and the reporting of grades, the value of effective instruction can be undone if the actual test questions are poorly constructed, unclear, or subjectively scored. Such problems do not give students a fair chance to show what they have learned and, consequently, do not provide a valid basis for decision making. No matter whether one is concerned with teacher-made, textbook, Ministry- or Department-mandated, or standardized tests, it is important that teachers be able to differentiate between well-constructed and poorly constructed test questions.

Tests are composed of short communications called questions or **items**. Each question must be brief and must set a clear problem for the student to think about. Each question must also be complete in itself and independent of other questions. Further, because students will mentally debate the nuances of each word to be sure they are not misinterpreting the intent of the item, it is crucial that questions be stated in clear, precise language. This chapter examines and contrasts different types of selected and constructed response test questions and provides general guidelines for writing or judging the adequacy of each kind of item. Let's begin with an introduction to summative assessment—assessment that takes place after learning has occurred.

## SUMMATIVE ASSESSMENT

In Chapters 2 and 4 we talked about diagnostic and formative assessment, respectively. Although critical to teachers' decision making, these forms of assessment should be supplemented by another form of assessment, **summative assessment**, which usually comes at the end of a classroom process, activity, unit, or course. Summative assessment is used mainly to

assess the outcomes of instruction and is exemplified by end-of-unit tests, projects, term papers, and final examinations. Table 5.1 contrasts diagnostic, formative, and summative assessments.

Summative assessments help teachers to make decisions that the school bureaucracy requires of them: testing, grading, and grouping of students; recommending whether students should be promoted or placed in special classes; and referring students to special education services if they have special needs. Because they have important consequences for students, summative assessments are generally based on systematically gathered evidence. They almost always focus on students' cognitive performance, usually how well students have learned what has been taught.

It is important to remember that the primary aim of summative assessment is to *provide students with a fair opportunity to demonstrate what they have learned from the instruction they have received*. The primary aim is not to trick students into doing poorly, entertain them, or ensure that most of them get A's. It is not to determine the total knowledge students have accumulated as a result of their learning experiences, both in and out of school. It is simply to let students show what they have learned from the things they have been taught in their classroom.

**TABLE 5.1** Comparison Of Diagnostic, Formative, And Summative Assessments

	<b>Diagnostic</b>	<b>Formative</b>	<b>Summative</b>
Purpose	To determine the ways in which teachers plan, teach, assess, and evaluate	To monitor and guide instruction while it is still in progress	To judge the success of instruction at its completion
Times of assessment	Beginning of instruction	During instruction	After instruction
Type of assessment technique	Observations, school records, discussions, comments	Observations, questions, seatwork, quizzes, homework	Tests, projects, papers, examinations
Use of assessment information	Hone the content and the method of instruction	Improve and change instruction while it is still going on	Judge the overall success of instruction; grade, place, promote

# SELECTED AND CONSTRUCTED RESPONSE TEST ITEMS

*Multiple-choice, true-false, and matching questions are examples of selected response items. Constructed response items are those in which the student constructs his or her own answer.*

There are two basic types of test questions: **selected response items** and **constructed response items**. As their names suggest, selected response items require the student to select the correct answer from among a number of choices, while constructed response items require the student to supply or construct his or her own answer.

## *Selected Response Items*

Within the general category of selected response items are multiple-choice, true-false, and matching questions.

### *Multiple-Choice Items*

*Multiple-choice items consist of a stem, which presents the problem or question, and a set of options from which the student selects an answer.*

Multiple-choice items consist of a **stem**, which presents the problem or question to the student, and a set of **options** from which the student selects an answer. The multiple-choice format is widely used in summative tests of all types, primarily to assess learning outcomes at the factual knowledge and comprehension levels. However, with suitable introductory material, this format can also be used to assess higher-level thinking involving application, analysis, and synthesis. The main limitations of the multiple-choice format are that it does not allow students to construct, organize, and present their own answers, and it is susceptible to guessing. However, these concerns about the use of multiple-choice items can be minimized through careful and considered item preparation (Cranton, 2000).

Here are two examples of multiple-choice items.

1. You use me to cover rips and tears. I am made of cloth. What am I?  
A. knot    B. patch    C. perch    D. scratch
2. What is the smallest province in Canada?  
A. New Brunswick  
B. Newfoundland  
C. Nova Scotia  
D. Prince Edward Island

Most often, multiple-choice questions do not effectively test higher-level thinking, but multiple-choice interpretive questions can do so. We will discuss them later in the section "Higher-Level Questions."

### True-False Items

The true-false format requires students to classify a statement into one of two categories: true or false; yes or no; correct or incorrect; fact or opinion. True-false items are used mainly to assess factual knowledge and comprehension behaviours, although they also can be used to assess higher-level ones (Frisbie, 1992). The main limitation of true-false questions is their susceptibility to guessing.

*The main limitation of true-false questions is their susceptibility to guessing.*

The following are typical true-false items.

*Although primarily used to assess knowledge and comprehension, both multiple-choice and true-false items can be used to assess higher-level thinking.*

1.  $5 + 4 = 8$  T F
2. In the equation  $E = mc^2$ , when  $m$  increases  $E$  also increases. T F
3. Read the statement. Circle T if true and F if false. If the statement is false, rewrite it to make it true by changing only the underlined part of the statement.

The level of the cognitive taxonomy that describes recall and memory behaviours is called the synthesis level. T F

### Matching Items

Matching items consist of a column of **premises**, a column of **responses**, and directions for matching the two. The matching exercise is similar to a set of multiple-choice items, except that in a matching question, the same set of options or responses is used for all the premises. Its chief disadvantage is that it is limited mainly to assessing lower-level behaviours. The following is an example of a matching exercise.

*Matching items consist of a column of premises, a column of responses, and directions for matching the two. They assess mainly lower-level thinking.*

On the line to the left of each invention in column A, write the *letter* of the person in column B who invented it. Each name in column B may be used only once or not at all.

#### Column A

- \_\_\_\_\_ (1) basketball
- \_\_\_\_\_ (2) pacemaker
- \_\_\_\_\_ (3) snowmobile
- \_\_\_\_\_ (4) telephone

#### Column B

- A. Alexander Graham Bell
- B. Frederick Banting
- C. James Naismith
- D. Joseph-Armand Bombardier
- E. Wilfred Bigelow

### Constructed Response Items

Constructed response items consist of short-answer and completion (also called fill-in-the-blank) items, essay questions, or questions requiring the student to create things such as diagrams or concept maps.

### *Short-Answer and Completion Items*

*Short-answer items use a direct question to present a problem; completion items use an incomplete sentence. Both tend to assess mainly factual knowledge and comprehension.*

Short-answer and completion items are very similar. Each presents the student with a question to answer. The short-answer format presents the problem with a direct question (e.g., What is the name of the first prime minister of Canada?), while the completion format may present the problem as an incomplete sentence (e.g., The name of the first prime minister of Canada is \_\_\_\_\_) or a picture, map, or diagram that requires labelling. In each case, the student must supply his or her own answer. Typically, the student is asked to reply with a word, phrase, number, or sentence, rather than with a more extended response. Short-answer questions are fairly easy to construct and diminish the likelihood that students will guess answers. However, they tend to assess mainly factual knowledge or comprehension.

The following are examples of completion and short-answer items.

1. Scientists who specialize in the study of plants are called \_\_\_\_\_.

Next to each province write the name of its capital city.

2. British Columbia \_\_\_\_\_
3. New Brunswick \_\_\_\_\_
4. Ontario \_\_\_\_\_
5. In a single sentence, state one way that inflation lowers consumers' purchasing power.

### *Essay Items*

*Essay questions are most useful for assessing higher-level thinking skills but are time-consuming to answer and score and favour the student with writing ability.*

Essay questions give students the greatest opportunity to construct their own responses, making them the most useful for assessing higher-level thinking processes like analyzing, synthesizing, and evaluating. The essay question is also the primary way teachers assess students' ability to organize, express, and defend ideas. The main limitations of essays are that they are time-consuming to answer and score, permit testing only of a limited amount of students' learning, and place a premium on writing ability.

Here are some examples of essay questions.

1. What is the value of studying science? Give your answer in complete, correct sentences. Write at least five sentences.
2. "In order for revolutionary governments to build and maintain their power, they must control the educational system." Discuss this statement using your knowledge of the American, French, and Russian revolutions. Do you agree with the statement as it applies to the revolutionary governments in the three countries? Include specific examples to support your conclusion. Your answer will be judged on the basis of the similarities and differences you identify in the three revolutions and the extent to which your conclusion is supported by specific examples. You will have 40 minutes to complete your essay.

## Comparing Selected and Constructed Response Items

Constructed response questions are much more useful than selected response questions in assessing students' ability to organize thoughts, present logical arguments, defend positions, and integrate ideas. Selected response questions, on the other hand, are more useful when assessing application and problem-solving skills. Given these differences, it is not surprising that knowing the kind of item that will be on a test can influence the way students prepare for the test. In general, constructed response items encourage global, integrative study, while selected response items encourage a more detailed, specific focus.

*Constructed response questions are most useful for assessing students' ability to organize and present their thoughts, defend positions, and integrate ideas.*

*Selected response items are most useful when application and problem-solving skills are assessed.*

**TABLE 5.2** Comparison of Selected and Constructed Response Test Items

	<b>Selected Response Items</b>	<b>Constructed Response Items</b>
<b>Types of Items</b>	Multiple-choice, true-false, matching, interpretive exercise	Short-answer, essay, completion
<b>Behaviours Assessed</b>	Factual knowledge and comprehension; thinking and reasoning behaviours like application and analysis when using interpretive exercises	Factual knowledge and comprehension; thinking and reasoning behaviours like organizing ideas, defending positions, and integrating points
<b>Major Advantages</b>	<ol style="list-style-type: none"> <li>1. Items can be answered quickly so a broad sample of instructional topics can be surveyed on a test.</li> <li>2. Items are easy and objective to score.</li> <li>3. Test constructor has complete control over the stem and options so the effect of writing ability is controlled.</li> </ol>	<ol style="list-style-type: none"> <li>1. Preparation of items is relatively easy; only a few questions are needed.</li> <li>2. Affords students a chance to construct their own answers; only way to test behaviours such as organizing and expressing information.</li> <li>3. Lessens chance the students can guess the correct answer to items.</li> </ol>
<b>Major Disadvantages</b>	<ol style="list-style-type: none"> <li>1. Time-consuming to construct.</li> <li>2. Many items must be constructed.</li> <li>3. Guessing is a problem.</li> </ol>	<ol style="list-style-type: none"> <li>1. Time-consuming to score.</li> <li>2. Covers small sample of instructional topics.</li> <li>3. Bluffing is a problem.</li> </ol>

While constructed and selected response items consume approximately the same amount of time to construct and score, each format allocates its time differently. Selected response items are time-consuming to construct, but can be scored quickly. Constructed response items are less time-consuming to construct, but are more time-consuming to score. Table 5.2 on page 93 summarizes the differences.

## HIGHER-LEVEL QUESTIONS

There is a growing emphasis on teaching and assessing students' higher-level thinking. As the following quotes show, teachers recognize the importance of students' learning how to understand and apply their knowledge. They know that knowledge takes on added meaning when it can be used in real-life situations.

Facts are important for students to learn in all subjects, but if students do not learn how to understand and use the facts to help them solve new problems, they haven't really learned the most important part of instruction.

The kids need to go beyond facts and rote learning. You can't survive in society unless you can understand, think, reason, and apply what you know.

It would be so boring to only teach facts. Some recall or memorization is needed, of course, but day after day of memorization instruction would be demeaning to my students and me. I have to make room in my curriculum for more complex thinking and reasoning skills such as understanding and applying new knowledge.

Many people believe that the only way to test higher-level thinking skills is with essay items. That is not the case! Any test question that demands more from a student than memory is a higher-level item. Thus, any item that requires the student to solve a problem, interpret a chart, explain something in his or her own words, or identify the relationship between two phenomena qualifies as an item of higher-level thinking. Similarly, any assessment that requires students to demonstrate their ability to carry out an activity (e.g., give an oral talk, construct a mobile, or read an unfamiliar foreign language passage aloud) also qualifies as being higher level.

### *Essay Questions*

Essay questions provide an important tool to assess higher-level thinking. Good essay questions require students to organize, understand, apply, integrate, and defend ideas. Questions that can be answered using only factual knowledge are better tested by more structured item types such as multiple-choice, true-false, or completion. The following are examples of essay questions that can elicit higher-level thinking from students. In all cases it



is assumed that the students have been taught material similar, though not identical, to that in the items.

1. Explain whether the reasoning in the following statements is correct or incorrect.  
     All dogs have tails.  
     This animal has a tail.  
     Therefore, this animal is a dog.
2. In what ways were the events leading up to the start of World War I the same as the events leading up to the start of World War II? In what ways were they different? Focus your answer on military, social, and economic factors.
3. Describe in your own words how an eclipse of the sun happens.
4. Why are some parts of the world covered by forests, some parts by water, some parts by grasses, and some parts by sand? Tell about some of the things that make a place a forest, an ocean, a grassland, or a desert.

### *Interpretive Exercises*

The interpretive exercise is a common form of multiple-choice item that can assess higher-level thinking. An interpretive exercise gives students some information or data and then asks a series of selection-type questions based on that information. The following is an example of an interpretive exercise:

*Interpretive exercises assess higher-level skills because the students must interpret or apply given information.*

Read the following passage.

- (1) For what men say is that, if I am really just and am not also thought just, profit there is none, but the pain and the loss on the
- (3) other hand is unmistakable. But if, though unjust, I acquire the reputation of justice, a heavenly life is promised to me. Since then
- (5) appearance tyrannizes over truth and is lord of happiness, to appearance I must devote myself. I will describe around me a
- (7) picture and shadow of virtue to be the vestibule and exterior of my house; behind I will trail the subtle and crafty fox.

Which one of the following states the major premise of the passage?

- A. For what men say (line 1)
- B. if I am really just (line 1)
- C. profit there is none, but the pain and the loss (line 2)
- D. appearance tyrannizes over truth and is lord of happiness (line 5)
- E. picture and shadow of virtue to be the vestibule and exterior of my house (line 7)

Figure 5.1 contains another example, taken from a sample question bank used for large-scale assessment by the Ministry of Education in Alberta. Generally, multiple-choice items that ask for interpretations of graphs, charts, reading passages, pictures, or tables are classified as interpretive exercises. Such exercises can assess higher-level behaviours like recognizing the relevance of information, identifying warranted and unwarranted generalizations, recognizing assumptions, interpreting experimental findings, and explaining pictorial materials.

To answer the questions posed, students have to interpret, comprehend, analyze, apply, or synthesize the information presented. Interpretive exercises assess higher-level skills because they contain all the information needed to answer the questions posed. Thus, if a student answers incorrectly, it is because he or she cannot do the thinking or reasoning required by the question, not because the student failed to memorize background information.

Like the essay question, the interpretive exercise is a useful way to assess higher-level thinking. However, unlike the essay question, interpretive exercises cannot show how students organize their ideas when solving a problem or how well they can produce their own answers to questions.

Other disadvantages of interpretive exercises are the difficulty of constructing them and the heavy reliance they often place on reading ability. Students who read quickly and with good comprehension have an obvious

**FIGURE 5.1**  
*Example of an Interpretive Exercise.*

Figure A

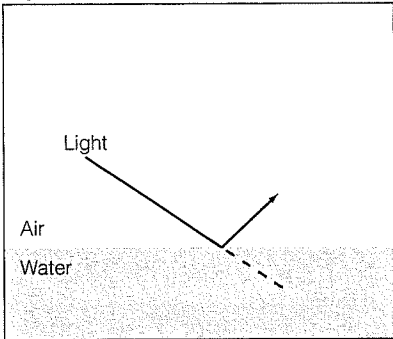
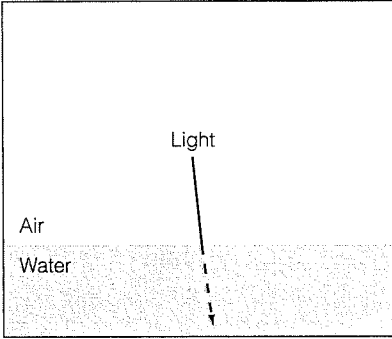


Figure B



1. When the angle of light rays to the surface of water is small—under 45 degrees—a high percentage of them are reflected (Figure A). When the angle of light rays to the surface of water is large—near 90 degrees—much of the light passes through the surface and enters the water (Figure B).

According to the information above, when would the greatest reflection of sunlight from the surface of a lake be observed?

- (1) late in the morning
- (2) when the sky is very cloudy
- (3) early in the afternoon
- (4) noon
- (5) late in the afternoon

advantage over students who do not. This advantage is particularly evident when the test involves reading and interpreting many passages in a limited amount of time.

Most teachers do not construct their own interpretive exercises, preferring to use exercises supplied by textbook publishers or other sources. Regardless of whether one is constructing or selecting an interpretive exercise, the exercise should meet five general guidelines before it is used to assess student achievement.

1. *Relevance.* The exercise should be related to the instruction provided students.
2. *Similarity.* The material presented in the exercise should be new to the students, but similar to material presented during instruction.
3. *Brevity.* There should be sufficient information for students to answer the questions, but the exercises should not become tests of reading speed and accuracy.
4. *Answers not provided.* The correct answers should not be found directly in the material presented. Interpretation, application, analysis, and comprehension should be needed to determine correct answers.
5. *Multiple questions.* Each interpretive exercise should include more than one question to make most efficient use of time.

Table 5.3 summarizes the pros and cons of the different types of selected and constructed response test items.

## GUIDELINES FOR WRITING AND CRITIQUING TEST ITEMS

Whether writing test items or selecting those prepared by others, there are three general guidelines that help insure good tests: (1) cover important objectives; (2) write clearly and simply; and (3) review items before testing. This section discusses and illustrates these guidelines.

### *Cover Important Objectives*

Note that the following discussion pertains to teacher-prepared tests and to textbook tests. Simply because a ready-made test is available with a textbook is no reason for a teacher to assume that the test adequately assesses his or her instruction on the chapter or unit. Each classroom teacher has a responsibility to decide about the suitability of a textbook test for assessing his or her instructional emphases.

*Test items should reflect important topics and skills emphasized during instruction, should be stated briefly and clearly, and should be self-contained.*

TABLE 5.3 Advantages and Disadvantages of Types of Test Items

Test Type	Advantages	Disadvantages
Multiple-choice items	<ol style="list-style-type: none"> <li>1. Large number of items can be given in a short period</li> <li>2. Higher- and lower-level objectives can be assessed</li> <li>3. Scoring is usually quick and objective</li> <li>4. Less influenced by guessing</li> </ol>	<ol style="list-style-type: none"> <li>1. Takes substantial time to construct items</li> <li>2. Not useful when "show your work" is required</li> <li>3. Often hard to find suitable options</li> <li>4. Reading ability can influence student performance</li> </ol>
True-false items	<ol style="list-style-type: none"> <li>1. A large number of items can be given in a short time</li> <li>2. Scoring is usually quick and objective</li> </ol>	<ol style="list-style-type: none"> <li>1. Guessing correct answer is a problem</li> <li>2. Difficult to find statements that are clearly true or false</li> <li>3. Items tend to stress recall</li> </ol>
Matching items	<ol style="list-style-type: none"> <li>1. An efficient way to obtain a great deal of information</li> <li>2. Easy to construct</li> <li>3. Scoring is usually quick and objective</li> </ol>	<ol style="list-style-type: none"> <li>1. Focus is mainly on lower-level outcomes</li> <li>2. Homogeneous topics are required</li> </ol>
Short-answer items	<ol style="list-style-type: none"> <li>1. Guessing is reduced; student must construct an answer</li> <li>2. Easy-to-write items</li> <li>3. Broad range of knowledge can be assessed</li> </ol>	<ol style="list-style-type: none"> <li>1. Scoring can be time-consuming</li> <li>2. Not useful for complex or extended outcomes</li> </ol>
Essay items	<ol style="list-style-type: none"> <li>1. Directly assess complex higher-level outcomes</li> <li>2. Take less time to construct than other item types</li> <li>3. Assess integrative, holistic outcomes</li> </ol>	<ol style="list-style-type: none"> <li>1. Difficult and time-consuming to score</li> <li>2. Provide a deep but small sample of students' performance</li> <li>3. Bluffing and the quality of writing can influence scores</li> </ol>
Interpretive exercise items	<ol style="list-style-type: none"> <li>1. Assess integrative and interpretive outcomes</li> <li>2. Assess higher-level outcomes</li> <li>3. Scoring is usually quick and objective</li> </ol>	<ol style="list-style-type: none"> <li>1. Heavily dependent on students' reading ability</li> <li>2. Difficult to construct items</li> </ol>

One important guideline to keep in mind when preparing tests is not to focus exclusively on trivial knowledge and skills. Studies that examined the nature of the test items written by classroom teachers have found that the vast majority assessed memory-level behaviours (Marso & Pigge, 1989,

1991). From elementary school to the university, items that stress recall and memory are used much more extensively than items that assess higher-level thinking and reasoning, mainly because it is much easier to write short-answer or multiple-choice questions. In far too many instances the richness of instruction is undermined by the use of test items that trivialize the breadth and depth of the concepts and skills taught.

Each example that follows states the objective taught, the test item used to assess it, and an alternative item that would have provided a more suitable assessment of the objective. Note that the poor items trivialized higher-level objectives by assessing them with a memory item.

1. **Objective:** Given a description of a literary form, the students can classify the form as fable, mystery, folktale, or fantasy.

**Poor item:** What kind of stories did Aesop tell? \_\_\_\_\_.

A. fables    B. mysteries    C. folktales    D. fantasies

**Better item:** A story tells about the year 2020 and the adventures of a young Martian named Zik, who travelled to other worlds to capture strange creatures for the zoo at Martian City. This story is best classified as a \_\_\_\_\_.

A. fable    B. mystery    C. folktale    D. fantasy

2. **Objective:** The students describe similarities and differences in chemical compounds and elements.

**Poor item:** Chlorine and bromine are both members of a chemical group called the \_\_\_\_\_.

**Better item:** Chlorine and bromine are both halogens. What similarities do they possess that make them halogens? What are two differences in their chemical properties?

There are two main reasons for ensuring that the questions in a test align with the important topics and skills that were emphasized during instruction. First, if there is not a good alignment between instruction and the test questions, performance on the test will be a poor indication of actual learning. Students may have learned what was taught but were stymied by an invalid test. Low grades usually accompany such invalid tests and can diminish students' effort and confidence.

Second, tests that do not align with instruction have little positive influence on motivating and focusing student study. If students find little relationship between instruction and test content, they will undervalue instruction. You can remember instances when you prepared well for a test based on the teacher's instruction and review only to find that the test contained many questions that focused either on picky, isolated details or on types of problems that were not discussed in class. Recall how you felt when you tried to prepare for the next test given by that teacher.

The problem of mismatch between tests and instruction can be overcome to a large degree by thinking about testing earlier than the day before the test is to be given. With relatively little advance planning, tests that assess the important aspects of instruction can be prepared.

*Test items that do not reflect the important topics of instruction are not valid indicators of student achievement.*

## *Write Clearly and Simply: Seven Rules*

If test questions use ambiguous words or sentence structure, include inappropriate vocabulary, or contain clues to the correct answers, the test will not be a valid indicator of student achievement. The most important skill in writing or selecting good test items is the ability to express oneself clearly and succinctly. Test items should be: (1) briefly stated so students do not spend a disproportionate amount of time reading; (2) clearly expressed so students understand their task; and (3) capable of standing alone since each item provides a separate measurement.

Following are seven rules for writing sound test items. Each is illustrated by some confusing test items prepared by teachers who knew the content they wanted to test but who were unable to clearly state their intent. A better version of the same items is also shown for comparison.

**Rule 1: Avoid ambiguous and confusing wording and sentence structure.** Students must understand test questions. If the wording or sentence structure is confusing and prevents students from figuring out what they are being asked, students cannot demonstrate their learning. Consider the following test items.

1. All but one of the following is not an element. Which one is not?  
A. carbon    B. plastic    C. salt    D. sugar
2. Newfoundland is not the only province that does not have a border with another province. T   F

In these examples, the wording and sentence construction are awkward and confusing. The student has to sort through multiple negatives to figure out what is being asked. It is better, therefore, to phrase questions briefly, directly, and in the positive voice, as shown in these edited versions.

1. Which one of these is an element?  
A. carbon    B. plastic    C. salt    D. sugar
2. Newfoundland is the only island province in Canada. T   F

Other items, such as examples 3 and 4, are more than just confusing, they are virtually incomprehensible.

3. What is the relative length of the shortest distance between Calgary and Winnipeg and Regina? \_\_\_\_\_
4. The \_\_\_\_\_ produced by the \_\_\_\_\_ is used by the green \_\_\_\_\_ to change \_\_\_\_\_ and \_\_\_\_\_ into \_\_\_\_\_. This process is known as \_\_\_\_\_.

What is a reasonable answer to each? Taken individually, the words in example 3 are not overly difficult, but their sequencing makes their intent altogether unclear. Example 4 is so mutilated with blank spaces that a student has to be a mind reader to figure out what is being asked. No student should be confronted by such a question. Students will answer items like

examples 3 and 4 incorrectly regardless of how well they have mastered the information and skills taught them. The following changes overcome the problems in these two examples.

3. Which is closer to Regina, Calgary or Winnipeg? \_\_\_\_\_
4. The process in which green plants use the sun's energy to turn water and carbon dioxide into food is called \_\_\_\_\_.

*Test items should be brief, clearly written, and free of ambiguous words so that comprehension is not an issue.*

If a student answers the revised items incorrectly, it is because he or she does not know the desired answer. That is acceptable. Remember, the purpose of a test item is not to guarantee correct answers, but to give students a *fair* chance to show how much they know about the things they were taught. To do this, test items must be readily comprehended.

Another factor that prevents students from being able to focus quickly and clearly on the question being posed is the use of ambiguous words or phrases. Read examples 5, 6, and 7 and try to identify a problem in each that could cause students difficulty in deciding how to answer.

5. Shakespeare was the world's greatest playwright. T F
6. The most important coastal city in Canada is \_\_\_\_\_.  
A. Halifax    B. St. John's    C. Vancouver    D. Victoria
7. Write an essay in which you consider the future of atomic energy.

Each example contains an ambiguous term that could be puzzling to students and make their choice of an answer difficult. The true-false example contains the undefined word *greatest*. Did the teacher mean that Shakespeare wrote more plays than any other playwright? that more of his plays are still being performed than those of any other playwright? that his plays are required reading in more Canadian classrooms than those of any other playwright? Until students know what the teacher means by *greatest*, they will have difficulty responding. Example 6 has the same fault. What does the phrase *most important* mean? Each of these cities is important in many ways. Words like *greatest*, *most important*, *best*, and similar ambiguous words should be replaced by more specific language, regardless of the type of test item used.

Note the rewritten versions of examples 5 and 6.

5. More of William Shakespeare's plays are required reading in Canadian classrooms than those of any other playwright. T F
6. The busiest port in Canada is \_\_\_\_\_.  
A. Halifax    B. St. John's    C. Vancouver    D. Victoria

In example 7 the teacher wants the students to consider the future of atomic energy. Does the teacher mean compare and contrast atomic energy to fossil fuel; discuss the relative merits of fission versus fusion as a means of generating energy; or explain the positive and negative consequences of increased use of atomic energy? It is not clear. The item needs

to be more specific for the students to respond in the way the teacher desires as shown in this revised version.

7. Describe the advantages and disadvantages of increased use of atomic energy in the automobile manufacturing process.

In most cases, the teachers who wrote the preceding examples knew what they wanted to ask students but were unable to write items that clearly conveyed their intent. Teachers must say precisely what they mean, not assume or hope that their students will interpret their test items in the ways intended.

**Rule 2: Use appropriate vocabulary.** The difficulty level of test questions can be influenced dramatically by vocabulary. If students cannot understand the vocabulary used in test questions, their test scores will reflect their vocabulary deficiencies rather than how much they have learned from instruction. Based on diagnostic assessments, every teacher should take into account the vocabulary level of his or her students when writing or selecting the items for achievement tests. Note the difference in the following two ways of writing a true-false question to assess students' understanding of capillary action, a principle that explains how liquids rise in narrow passages.

The postulation of capillary effectuation promotes elucidation of how pliant substances ascend in incommodious veins. T F

The principle of capillary action helps explain how liquids rise in small passages. T F

Clearly, vocabulary level can affect the ability of students to understand what is being asked in a test question.

**Rule 3: Keep questions short and to the point.** Items should quickly focus students on the question being asked. Examine these questions.

*Questions should be short, specific, and written at students' vocabulary level.*

8. Switzerland
  - A. is located in Asia.
  - B. produces large quantities of gold.
  - C. has no direct access to the ocean.
  - D. is a flat, arid plain.
9. Billy's mother wanted to bake an apple pie for his aunt and uncle, who were coming for a visit. Billy had not seen them for many months. When Billy's mother saw that she had no apples in the house, she sent Billy to the store to buy some. Her recipe called for 8 apples to make a pie. If apples at the store cost 30 cents for two, how much money will Billy need to buy eight apples?
 

A. \$.30    B. \$.90    C. \$1.20    D. \$2.40

In example 8, the stem does not clearly set a problem for the student, that is, after students read the item stem Switzerland, they still have no idea of the question being asked. Only after reading the stem *and* all the options



does the point of the item begin to become clear. The item could be more directly stated as follows.

8. Which of the following statements about the geography of Switzerland is true?
- A. It is located in Asia.
  - B. It is a flat, arid plain.
  - C. It has no direct access to the ocean.
  - D. It has a tropical climate.

Example 9 is intended to determine whether the student can correctly calculate the cost of some apples. The information about the aunt and uncle's visit, how long it had been since Billy last saw them, or the lack of apples in the house is not important, can be distracting, and takes time away from relevant items. A better way to state the item is shown here.

9. To make an apple pie Billy's mother needed 8 apples. If apples cost 30 cents for two, how much will 8 apples cost?
- A. \$.30    B. \$.90    C. \$1.20    D. \$2.40

In short-answer or completion items, the blanks should come at the end of the sentence so students know what kind of a response is required. Compare these two items and notice how placing the blank at the end helps convey what the item is about.

\_\_\_\_\_ and \_\_\_\_\_ are two Canadian First Nations languages.

Two Canadian First Nations languages are \_\_\_\_\_ and \_\_\_\_\_.

Matching items can also be written to help students focus more quickly on the questions being asked. Look over example 10 and suggest a change that would focus students more clearly on the questions they have to answer.

10. Draw a line to match the famous scientist in column A with his or her accomplishment in column B.

**Column A**

Albert Einstein  
Charles Darwin  
Marie Curie  
Sir Isaac Newton

**Column B**

derived the equation,  $E=mc^2$   
discovered radium  
invented a form of mathematics called calculus  
proposed that nebulas were composed of stars  
proposed the theory of natural selection

Most matching items can be improved by placing the column with the lengthier descriptions on the left and the column with the shorter descriptions on the right, as shown next.

10. Draw a line to match the famous scientist in column B with his or her accomplishment in column A.

**Column A**

Derived the equation,  $E=mc^2$   
 Discovered radium  
 Invented a form of mathematics called calculus  
 Proposed that nebulas were composed of stars  
 Proposed the theory of natural selection

**Column B**

Albert Einstein  
 Charles Darwin  
 Marie Curie  
 Sir Isaac Newton

*With the exception of essays, most test items should have only one correct answer.*

**Rule 4: Write items that have one correct answer.** With the exception of essay questions, most test items are designed to have students select or construct one best answer. With this goal in mind, read examples 11 and 12. See how many correct answers you can provide for each item.

11. Who was Sir John A. Macdonald? \_\_\_\_\_  
 12. Ernest Hemingway wrote \_\_\_\_\_.

Each of these items has more than one correct answer. Sir John A. Macdonald is most well known for being the first prime minister of Canada. But he was also the second longest serving prime minister, a father of Confederation, a member of the Conservative Party, a lawyer, and a Knight Commander of the Order of St. Michael and St. George. Faced with such an item, students ask themselves which of the many things I know about Sir John A. Macdonald should I answer? Similarly, Ernest Hemingway wrote short stories and letters, in Spain, in pencil, as well as famous novels such as *The Old Man and the Sea*.

Examples 11 and 12 should be restated so that students know precisely what is being asked. Notice how each question asks for something specific—a name—thus indicating to students the nature of the expected answer.

11. What is the name of the first prime minister of Canada? \_\_\_\_\_  
 12. The name of the author of *The Old Man and the Sea* is \_\_\_\_\_.

Items with more than one correct answer occur much more often in short-answer and completion items than in selected response items. Unless short-answer or completion items are stated specifically and narrowly, the teacher can expect many different responses. The dilemma for the teacher then becomes whether to give credit for answers that are technically correct but not the desired one.

**Rule 5: Give information about the nature of the desired answer.** While the failure to properly focus students is common to all types of test items, it is most often seen in essay items. Despite students' freedom to structure their own responses, essay questions should still require students to demonstrate mastery of key ideas, principles, or concepts that were taught. An essay, like any other type of test item, should be constructed to find out how well students have learned the things they were taught.

Here are two typical essay questions written by classroom teachers.

13. Describe what happened to art during the Renaissance.
14. Why should you study science?

In each of these questions, the students' task is not clearly defined. When students encounter global questions such as these they may have little idea of what the teacher is looking for and may end up with a poor grade because they incorrectly guessed the teacher's intent. This practice is unfair to students and produces test results that do not reflect their achievement.

To determine whether students have learned what was taught, essay questions should be narrowed to focus students on the areas of interest. Students should be informed about the nature and scope of the expected answer. While essay questions should provide the student freedom to select, organize, state, and defend positions, they should not give students total freedom to write whatever they want. Obviously, to develop a well-focused essay question the teacher must give considerable thought to the purpose and scope of the question before actually writing it.

Examples 13 and 14 and have been rewritten to more precisely reflect the teacher's intent. Notice how the vague and ambiguous directions are made clearer to students in the revised questions.

13. Compare art during the Renaissance to art prior to the movement in terms of the portrayal of the human figure, use of colour, and emphasis on religious themes. Your essay will be judged in terms of the distinctions you identify between the two periods and the explanations you provide to account for the differences.
14. Give two reasons for why we should study science. What are some things that studying science teaches us? What are some jobs that use science? Write your answer in at least five complete sentences.

Certainly these are not the only ways that these essay items could have been rewritten, but these revisions point out the need for focus in essay questions. When students approach these revised items, they have a clear sense of what is expected of them; they no longer have to guess what the scope and direction of their answers should be. Note also that it is much more difficult for the student to bluff an answer to the revised items than it is to the initial, broadly stated items. The revised items call for answers specifically related to instruction, and therefore test what was taught and make scoring easier.

To summarize, regardless of the particular type of test item used, students should be given a clear idea of what their task is. In the case of multiple-choice items this may mean elaborating a stem in order to clarify the options. In matching items it may involve putting the longer options in the left column. In short-answer or completion items it may mean placing the blank at the end of the statement or specifying precisely the

*Essay questions should focus students' answers on the major points covered by instruction.*

*For all types of test items, students should have a clear sense of what is expected of them.*

*Test item writers should take care not to provide grammatical clues, implausible option clues, sequencing clues, sentence length clues, or specific determiner clues.*

nature of the desired answer. In essay questions it may mean elaborating to include information about the scope, direction, and scoring criteria for a desired answer. In all cases, the intent is to allow the student to respond validly and efficiently to the items.

**Rule 6: Do not provide clues to the correct answer.** The item-writing rules discussed thus far have all been aimed at problems that inhibited students from doing their best. However, the opposite problem arises when test items contain clues that help students answer questions correctly even though they have not learned the content being tested. Many types of clues may appear in items: grammatical clues, implausible option clues, sequencing clues, sentence length clues, and specific determiner clues. Try to identify the clue in examples 15 and 16.

15. A figure that has eight sides is called an \_\_\_\_\_.
- A. pentagon    B. quadrilateral    C. octagon    D. ogive
16. Compared to autos of the 1960s, autos in the 1980s \_\_\_\_\_.
- A. more horsepower.  
B. to use more fuel.  
C. contain more safety features.  
D. was less often constructed in foreign countries.

These examples contain grammatical clues. In example 15, using the article *a* or *an* at the end of the question or stem indicates to students what letter will begin the next word. The *an* before the blank tells the student that the next word must begin with a vowel, so the options pentagon and quadrilateral cannot be correct. There are two ways to correct this problem: replace the single article with the combined *a(n)* or get rid of the article altogether by writing the question in the plural form.

15. Figures that have eight sides are called \_\_\_\_\_.
- A. octagons    B. ogives    C. pentagons    D. quadrilaterals

In example 16, only option C grammatically fits the stem. Regardless of students' knowledge, they can select the correct answer because of the grammatical clue. The corrected item might read:

16. Compared to autos of the 1960s autos in the 1980s \_\_\_\_\_.
- A. are always constructed in foreign countries.  
B. contain more safety features.  
C. have more horsepower.  
D. use more fuel.

Now try to find the clues in examples 17 and 18.

17. Which of the following best describes an electron?
- A. a voting machine  
B. negative particle  
C. neutral particle  
D. positive particle

18. Match the correct phrase in column A with the term in column B. Write the *letter* of the term in column B on the line in front of the correct phrase in column A.

**Column A**

- \_\_\_\_\_ 1. type of flower  
 \_\_\_\_\_ 2. poisonous snake  
 \_\_\_\_\_ 3. how amoebae reproduce  
 \_\_\_\_\_ 4. colour of chlorophyll  
 \_\_\_\_\_ 5. chemical element

**Column B**

- A. cobra  
 B. fission  
 C. green  
 D. hydrogen  
 E. rose

Example 17 contains a clue that is less obvious than those in examples 15 and 16, but is quite common in multiple-choice items. One of the options is inappropriate or implausible and therefore is immediately dismissed by the students. Choice A, a voting machine, is dismissed as an unlikely answer by all but the most careless readers. As much as possible, options in test questions should be realistic and reasonable. A useful strategy is to have at least three incorrect (but reasonable) options, or **distractors**, in each multiple-choice item.

*A distractor is a reasonable but incorrect option in a multiple-choice item.*

The more choices students have, the less likely it is that they can guess the correct answer. Understanding this, teachers sometimes write three or four good options for an item and then add a fourth or fifth, such as none of the above or all of the above. It is usually better to avoid such general options.

Example 18 is a very easy question; the topics are so different from one another that many of the options in column B are implausible matches to the statements in column A. The item does not test one homogeneous subject area.

Consider the following matching item, that tests students' knowledge of a single, homogeneous topic. Note the difficulty in answering this item compared to the previous version of example 18.

*A matching item should test the students' knowledge of a single homogeneous topic.*

18. Match the names of the animals in column A to their correct classification in column B. Write the *letter* of the correct classification on the line in front of each animal name. The choices in column B may be used more than once.

**Column A**

- \_\_\_\_\_ 1. alligator  
 \_\_\_\_\_ 2. condor  
 \_\_\_\_\_ 3. frog  
 \_\_\_\_\_ 4. porpoise  
 \_\_\_\_\_ 5. snake  
 \_\_\_\_\_ 6. salamander

**Column B**

- A. amphibian  
 B. bird  
 C. fish  
 D. mammal  
 E. reptile

The revised item is a better test of students' knowledge in two ways. First, it does not include the obvious matches and mismatches that occur when many unrelated topics are contained in the same matching item. The revised version focuses on a single topic, classification of animals into

groups. Second, unlike example 18, the revised item has an unequal number of entries in columns A and B. Unequal entries in the two columns of a matching item prevent students from getting the last match correct by the process of elimination.

Look for the clues in examples 19 and 20.

19. Some people think the moon is made of green cheese. T F
20. One should never phrase a test item in the negative. T F

These items contain clues that are called **specific determiners**. In true-false questions, words such as *always*, *never*, *all*, and *none* tend to appear in statements that are false, and testwise students tend to answer accordingly. These students often use ordering of options and their varying lengths as clues. Conversely, words like *some*, *sometimes*, and *may* tend to appear in statements that are true. Thus, in example 19, it is reasonable to assume that *some* people think the moon is made of green cheese, so T should be marked. On the other hand, example 20 must be marked F if there is even a single situation in which a test item can reasonably be stated in the negative (e.g., Which one of these is *not* an example of democracy?).

**Rule 7: Don't overcomplicate test items.** Occasionally, teachers and textbooks overcomplicate test items. Consider the following item, which was given to Grade 6 students to test their mastery of applying the procedure to calculate simple interest.

John borrowed \$117.55 from Bob at an interest rate of 9.73 percent a year. How much simple interest must John pay Bob at the end of 15 months?

The numbers in this example are difficult and almost ensure that many students in Grade 6 will make computational errors. Unless the teacher was specifically testing computational accuracy, the following example would better assess the students' ability to apply the procedure.

John borrowed \$150.00 from Bob at an interest rate of 9.00 percent a year. How much simple interest must John pay Bob at the end of one year?

The latter item assesses students' mastery of simple interest without complicating the computation so much that errors are likely to occur. Key Assessment Tools 5.1 summarizes advice regarding different types of items.

### *Review Items before Testing*

The best advice that can be given to improve most classroom tests is to review them before reproducing and administering them to students. Having written or selected the items for a chapter or unit test, it is recom-

*It is helpful to have a colleague or friend critique test items before the test is administered to students.*

mended that a teacher wait one day and then reread them. The teacher should also ask a colleague, spouse, or friend to review the items critically.

## *Decisions in Planning a Test*

Let us refer back to Rita's plan for instruction and assessment of learning outcomes from her Grade 5 unit on weather. She will need to answer four questions in her preparation of an end-of-unit test.

**1. What should I test?** The first important decision when preparing to assess student achievement is to identify the information, processes, and skills that will be tested. A valid assessment is one that provides students a fair opportunity to show what they have learned from instruction. Therefore, in deciding what to test, it is necessary for Rita to focus attention upon both her objectives and the actual instruction that takes place. Usually the two are very similar, but sometimes it is necessary to add or omit an outcome once teaching begins.

*A fair and valid test covers information and skills similar to those covered during instruction.*

### KEY ASSESSMENT TOOLS 5.1

#### Writing Test Items

##### Multiple-choice items

1. Put each test item on a different line.
2. Place the student tasks in the item stem.
3. Put repeated terms in the item stem.
4. Construct at least three alternative choices.
5. Put options in alphabetical order.
6. Avoid grammatical clues to the answer.
7. Be sure that items match students' reading level.
8. Do not include unneeded words.
9. If "no" is used, underline it.
10. Reread the item to identify spelling and other errors.
11. The item stem should clearly state the question to be answered.

##### True-false items

1. Make items clearly true or false.
2. Be sure that the item is important in the assessment.
3. Avoid specific determiners.
4. Make the true and false items about the same length.
5. Do not use items in a repetitive pattern.
6. Do not use textbook sentences.

*(continued)*

**KEY ASSESSMENT TOOLS 5.1****Writing Test Items (Continued)****Short-answer items**

1. Make sure the item relates to the assessment being taught.
2. Provide a clear focus for the intended answer.
3. Make answers possible in short responses; construct item so that student answers are short.
4. Ask students to reply in only one or two responses.
5. Be sure item assesses the intended responses.
6. Put space for the answer at the end of the item.
7. Avoid giving grammatical clues.

**Matching answer items**

1. Be sure the exercise relates to the assessment.
2. Compare homogeneous topics.
3. Make sure the directions are clear.
4. Put the longer options in the left-hand column.
5. Put options in alphabetical order.
6. Number one set of items; mark the other set with letters.
7. Do not ask for more than 10 responses in the assessment. If more responses are desired, begin a new matching test.
8. Provide one or two additional options in one column to avoid the final option being the correct answer by default.

**Essay questions**

1. Use several short essays rather than one long one.
2. Be sure that the reading level is appropriate for students.
3. Be sure the essay relates to your objectives.
4. Base the essay on a fresh example, if examples are used.
5. Provide a clear focus of the desired outcome of the essay.
6. Do not use essays that require a great deal of memory.
7. Aid students by focusing them with terms such as “state and defend the topic,” “apply the principle to,” “develop a valid conclusion,” and so on. These kinds of instructions focus students and also help to focus the grading.
8. Provide students with clear directions about the expected length of essay and the amount of time for completion.
9. Provide students with clear scoring criteria, for example, will spelling count?

Based on the expectations or learning outcomes from this Grade 5 unit, Rita knows that she has to determine how well the students are able to discriminate between *weather* and *climate*, predict weather patterns using data from weather reports, describe the water cycle, define terms such as *tem-*



perature, precipitation, humidity, wind chill factor, barometric pressure, and cloud cover, and describe ways in which weather conditions affect the activities of humans and animals. But what about other important skills such as designing, constructing, and assessing a variety of weather instruments? These are also covered in the unit, so should they be on Rita's test?

The answer to this question is no! There will always be more expectations or learning outcomes to teach than there is time to teach them. There will always be useful topics and skills that have to be omitted from tests because of a lack of instructional time. Including untaught skills on a test diminishes its validity, making it less than a true and fair assessment of what students have learned from classroom instruction. There also exist other formats such as projects, in-class assignments, portfolios, and performance assessments that can be used to assess those other important skills that are not easily assessed using written tests.

**2. What type of assessment should be given?** This question is answered by referring back to the expectations or learning outcomes. Each expectation or outcome contains a target process or behaviour that students have been taught. Knowledge outcomes should be tested using *selected response questions*, questions that present the student with a set of choices of which the student selects one or more. Outcomes that refer to comprehending, applying, analyzing, synthesizing, and evaluating may be tested using selected response questions, but are best assessed using *constructed response questions*, questions that require the student to construct an answer or product. Thus the statement of the outcome largely pre-determines the format used to assess learning. In any case, written tests primarily assess cognitive skills; affective and psychomotor skills are best assessed using other formats.

Many teachers feel that only essay tests are good. Others use multiple-choice items as much as possible, and still others believe that tests should contain a variety of question types. It must be remembered that the main purpose of summative assessment is *to permit students to show how well they have learned the knowledge, behaviours, or processes they were taught*. Thus, no single type of assessment item is applicable all the time. What makes a particular procedure useful is whether it matches the outcomes and instruction provided.

**3. How long should the test take?** Since time for testing is limited, choices must be made in deciding on the length of a test. Usually, practical matters such as the age of the students or the length of a class period are most influential. Since the stamina and attention span of young students is less than that of older ones, a useful strategy to follow with elementary school students is to test them fairly often using short tests that assess only a few outcomes. Because of their typical attention span, 15- to 30-minute tests, depending on the grade and group, are suggested. Testing in middle, junior, and high school is usually restricted by the length of the class period. Most teachers at these levels plan their tests to last almost one complete class period. For more on length of test time, see Chapter 6.

*The type of assessment procedure chosen depends on the nature of the outcome being assessed.*

*The age of the students, the subject being tested, and the length of the class period all impact the length of a test.*

*The number of test questions per outcome depends on the instructional time spent on each outcome and its importance.*

*The main consideration in judging the adequacy of a textbook test is the match between its questions and what students were actually taught in class.*

In deciding how many questions to ask for each outcome, try to balance two factors: (1) the instructional time spent on each outcome and (2) its importance. Some outcomes are usually more important than others. These outcomes tend to be more general ones that call for the integration of several specialized outcomes. However, if considerable time is spent teaching a specific outcome, it should be tested separately. When teachers focus their tests solely upon the general, integrative outcomes, students may answer questions incorrectly because they cannot successfully integrate the separate skills that they have learned.

**4. Should a teacher-made or a textbook test be used?** Teachers are inevitably confronted with the question of whether to use the textbook test or construct their own. The very availability of textbook tests is seductive and causes many teachers to think: the test comes with the textbook and seems to measure what is in the chapter that I'm teaching, so why shouldn't I use it?

The decision about using a textbook test or constructing one cannot be answered until *after* the teacher has reflected on what was taught and has identified the topics and behaviours to be tested. The usefulness of any test cannot be judged without reference to the planned outcomes and actual instruction.

Textbook tests furnish a ready-made instrument for assessing the outcomes stressed in the textbook and can save teachers much time. Test formats vary across textbook publishers in terms of length, layout, and question type. Look through the teacher's edition of some textbooks to see the range of tests available.

Regardless of whether a teacher is constructing his or her own test or judging the adequacy of a textbook test, he or she must consider the same basic validity issue: Do the items on the test match the instruction provided to the students? The more a teacher alters and reshapes the textbook curriculum, the less valid its accompanying tests become.

It is possible to combine textbook items and teacher-constructed items into an assessment. Often the textbook test has some appropriate assessment items that can be used in conjunction with the items the teacher has constructed. The key issue, however, is the relevance of the assessment items to the instruction provided to the students. Key Assessment Tools 5.2 relates to judging textbook tests.

To summarize, both textbook and teacher-made tests should (1) clearly relate to the outcomes of instruction, (2) include enough questions to assess all or most of the outcomes, and (3) use assessment methods suited to the backgrounds and prior experiences of the students (Joint Advisory Committee, 2002). Tests that meet these criteria will provide a valid indication of student learning. Key Assessment Tools 5.3 provides a summary of common problems teachers encounter in judging tests used for summative purposes.

## Assembling Tests

Once items have been written or selected and reviewed, they must be arranged into a test. If a teacher uses a textbook test, the items will already be arranged and ready for copying. Often teachers cut and paste items from various sources into a single test. In assembling a test, similar types of items should be grouped together and kept separate from other item types. For example, all of the short-answer questions should be together and separate from the multiple-choice, matching, completion, and essay questions. Grouping test items by type avoids the necessity of students shifting from one response mode to another as they move from item to item. It also means that a single set of directions can be used for all of the items in that test section, helping students cover more items in a given time. Finally, grouping test items makes scoring easier.

Another important consideration in assembling the test is the order in which the item types are presented. In most tests, selected response items come first and constructed response items come last. Within the constructed response section, short-answer or completion questions should be placed before essay questions. Constructed response items are placed at the end of the test so that students will not devote a disproportionate amount of time to this part of the test.

### KEY ASSESSMENT TOOLS 5.2

#### Key Points to Consider in Judging Textbook Tests

1. The decision to use a textbook test must come *after* a teacher identifies the outcomes that he or she has taught and wants to assess.
2. Textbook tests are designed for the generic classroom, but since few classrooms are generic, most teachers deviate somewhat from the text in order to accommodate their students' needs.
3. The more classroom instruction deviates from the textbook, the less valid the textbook tests are likely to be.
4. The main consideration in judging the adequacy of a textbook test is the match between its test questions and what students were taught in their classes:
  - a. Are questions similar to the teacher's objectives and instructional emphases?
  - b. Do questions require students to perform the behaviours they were taught?
  - c. Do questions cover all or most of the important outcomes taught?
  - d. Is the language level and terminology appropriate for the students?
  - e. Does the number of items for each outcome provide a sufficient sample of student performance?

**KEY ASSESSMENT TOOLS 5.3****Common Problems in Developing or Selecting Tests to Assess Student Achievement**

1. Failing to consider outcomes and instructional emphases when planning a test.
2. Failing to assess all of the important outcomes and instructional topics.
3. Failing to select item types that permit students to demonstrate the desired behaviour.
4. Adopting a test without reviewing it for relevance to the instruction provided.
5. Including topics or outcomes not taught to students.
6. Including too few items to assess the consistency of student performance.
7. Using tests to punish students for inattentiveness or acting out.

When arranging items on a test, remember these practices:

1. Designate a space for students to write their names and/or ID numbers (if applicable).
2. Do not split a multiple-choice or matching item across two pages of the test. This can cause unintended errors when students flip from one page to the next to read the second half of a matching question or the last two options of a multiple-choice question.
3. Separate multiple-choice options from the stem by beginning the options on a new line.
4. Number test items, especially if students must record answers on a separate answer sheet or in a special place on the test.
5. Space items for easy reading and provide enough space for students to complete constructed response items. Remember that young students write large and need space to respond. Do not cram items close together.

Each section of a test should have directions that focus students on what to do, how to respond, and where to place their answers. Lack of clear directions is one of the most common faults in teacher-prepared tests and often influences test validity. Here are some sample directions.

- Items 1–15 are multiple-choice items. Read each item carefully and write the letter of your answer on the line in front of the question number.
- Use words from the boxes to complete the sentences. Use each word only once.
- Answer each question by writing the correct answer in the space below the question. No answer should be longer than one sentence.

- For items 10 to 15, circle T or F (true or false).
- Use the chart to help you answer questions 27–33. Write your answers in the space provided after each question.

Directions such as these at the start of a test section focus students by telling them where and how to respond to the questions. To emphasize a point made earlier, it is especially important that each essay question spell out clearly for students the scope and characteristics of the desired answer. When applicable, it is also helpful to indicate the number of marks that will be given to each test section so they can make decisions about how to allocate their time.

The test should be reproduced so that each student has his or her own copy. Writing the test questions on the blackboard can be time-consuming, create problems for students with poor vision, and encourage students to look around the room during test taking. Orally reading questions can be used to pace students, but this approach places a premium on listening ability and prevents students from working at their own pace. This practice should be avoided unless one is assessing listening skills.

Tests that contribute to valid decisions also need to be reliable, that is, produce consistent scores. Without reliability, a test can hardly provide the kind of information on which a teacher would want to base decisions about students' learning. The main factors in attaining reliable summative tests are (1) the number and representativeness of the items included on the test and (2) the objectivity of scoring. In general, longer tests allow a teacher to look at a greater number of objectives. For example, which test do you think would produce the more stable and consistent information about a student's spelling achievement: a test that consists of a single word selected from a 100-word list or a test that contains a sample of 15 words selected from the same 100-word list? Key Assessment Tools 5.4 summarizes guidelines for assembling tests.

Finally, it is important that the assembled test be reviewed before reproducing and administering it. After constructing a test, it is recommended that a teacher wait one day and reread the test. If possible, the teacher should also ask a colleague to review the test format and individual test items critically.

## ACCOMMODATIONS: SUBSTITUTIONS FOR WRITTEN TESTS

There are a number of alternatives to written tests for students with exceptionalities. For example, the simplest way to accommodate a student who is physically unable to respond in writing to a test of subject knowledge and ideas is to present the test items orally. In some cases it may be appropri-

### TECHNOLOGY AND ASSESSMENT

<http://escholarship.bc.edu>

The *Journal of Technology, Learning and Assessment* (JTLA) is a peer-reviewed, scholarly online journal that produces interdisciplinary articles combining technology, learning theory, and assessment.

Find a link to this site and others on our OLC at

[www.mcgrawhill.ca/college/airasian](http://www.mcgrawhill.ca/college/airasian).

**KEY ASSESSMENT TOOLS 5.4****Guidelines for Assembling a Test**

This list combines some suggestions to consider when assembling a test:

- Organize the test by item type: selected response before constructed response, essay last.
- Indicate the purpose of the test on the top of the first page of the test.
- Allow sufficient space for written responses, especially for young students.
- Do not split multiple-choice or matching items across two pages.
- Separate stem from options in multiple-choice questions.
- Number test items.
- Provide clear directions for each section of the test; for older students, indicate the value of each section or question.
- Provide enough questions to ensure reliability.
- Proofread the test before copying.

ate for a student with a physical disability to use a computer in place of paper and pencil. In other cases, special equipment may be needed to make the computer itself accessible to such a student. For students with extreme physical limitations there may be workable accommodations, such as speech-to-text software for students unable to type. There may be a special education teacher or classroom in your school that can help with this type of problem, which is likely to require some forethought. Figure 5.2 is a suggestive rather than definitive list of products and Web site resources that can be useful for accommodating students with exceptionalities when assessing for summative purposes.

FIGURE 5.2 *Resources for Universal Design.*

<b>Software</b>	<b>Hardware</b>	Berkeley Digital Library SunSite
<b><i>Built-in Accessibility of Operating Systems</i></b>	<b><i>Portable Word Processors</i></b>	www.sunsite.berkeley.edu
Apple Special Needs www.apple.com (search for special needs)	AlphaSmart www.alphasmart.com	The Children's Literature Web Guide www.ucalgary.ca/~dkbrown
Microsoft Enable www.microsoft.com/enable	CalcuScribe www.calcuscribe.com	Internet Public Library www.ipl.org
<b><i>Text-to-Speech Software Programs</i></b>	DreamWriter www.brainium.com	Project Gutenberg www.gutenberg.org
CAST eReader www.cast.org Kurzweil 3000 www.kurzweilededu.com	LaserPC6 www.perfectsolutions.com	University of Virginia Library Electronic Text Center www.etext.lib.virginia.edu/ ebooks
ReadPlease 2003 www.readplease.com	QuickPad www.quickpad.com	<b>Organizations</b>
TextHELP! www.texthelp.com	<b><i>Handwriting Recognition Technologies</i></b>	<b><i>Technology in Education</i></b>
Write: OutLoud www.donjohnston.com	InkLink www.siibusinessproducts.com	Association for the Advancement of Computing in Education (AACE) www.aace.org
WYNN www.freedomscientific.com	InkWell www.apple.com/macrosx/ features/inkwell	Association for Educational Communications and Technology (AECT) www.aect.org
<b><i>Speech-to-Text Software Programs</i></b>	Logitech io Personal Digital Pen www.logitech.com	International Society for Technology in Education (STE) www.iste.org
Dragon Naturally Speaking www.scansoft.com IBM ViaVoice www.nuance.com/viavoice/ pro	PenReader www.smarttech.com	U.S. Department of Education Office of Educational Technology www.ed.gov/Technology
<b><i>Accessible Multimedia</i></b>	<b><i>Electronic Whiteboards</i></b>	<b><i>Accessibility</i></b>
HiSoftware www.hisoftware.com	Mimio www.mimio.com	CPB/WGBH National Center for Accessible Media www.ncam.wgbh.org
	SMARTBoard www.smarttech.com	
	<b>Online Resources</b>	
	<b><i>Digital Text</i></b>	
	American Library Association Great Sites for Children www.ala.org/parentspage/ greatsites/lit.html	

SOURCE: Curry, C. (2003). Universal design accessibility for all learners. *Educational Leadership*, (61) 2, 55-60.

## CHAPTER REVIEW



### CHAPTER REVIEW

Visit Chapter 5 of the Online Learning Centre at [www.mcgrawhill.ca/college/airasian](http://www.mcgrawhill.ca/college/airasian) to take chapter quizzes, link to related Web sites, and read PowerWeb articles and news feed updates.

#### *What are summative assessments?*

- Summative assessments are assessments that take place after learning has occurred in order to judge the success of the instruction. They usually take the form of tests, projects, papers, or examinations.
- Summative assessments are useful for student promotion and placement considerations.
- Written tests almost always focus on students' cognitive performance.

#### *What are the two basic types of written test items?*

- The two basic types of test items are selected response and constructed response.
- Selected response items are those in which the student selects the correct answer from among a number of choices. Examples of selected response items include multiple-choice items, true-false items, matching items, and interpretive exercises.
- Constructed response items are those that require the student to construct his or her own answer. Examples of constructed response items include short-answer items, completion items, and essay items.

#### *What are the parts of any multiple-choice item?*

- Multiple-choice items consist of a *stem*, which presents the problem or question, and a set of *options* from which the student selects an answer.
- Among the set of options, there is one correct option and, usually, at least three incorrect (but reasonable) options, or *distractors*.

#### *What are some rules for writing sound test items?*

- Avoid ambiguous and confusing wording and sentence structure.
- Use appropriate vocabulary.
- Keep questions short and to the point.
- Write items that have one correct answer.
- Give information about the nature of the desired answer.
- Do not provide clues to the correct answer.
- Do not overcomplicate test items.

#### *What are specific determiners?*

- Specific determiners are clues in true-false questions that assist the student in answering the question correctly.
- Examples of specific determiners are words such as *always*, *never*, *all*, and *none*, which tend to appear in statements that are false, and words like *some*, *sometimes*, and *may*, which tend to appear in statements that are true.



- Specific determiners are to be avoided in the construction of test items.

## QUESTIONS FOR DISCUSSION

1. What are some expectations or learning outcomes that are best assessed by selected response items? What are some expectations or learning outcomes that are best assessed by constructed response items?
2. How are diagnostic assessment and instruction related to written tests of student learning?
3. What are the pros and cons of giving students choices in answering essay items?
4. What harm could result if a teacher's tests produced invalid information about student learning?
5. Are higher-level expectations or learning outcomes harder to teach and assess than lower-level ones? Why or why not?
6. Why is the development of higher-level thinking in students important to society?
7. What are the specific challenges that children with exceptionalities face when working with written tests?

## ACTIVITIES

Each of the following five test items has at least one fault. Read each item, identify the fault(s) in it, and rewrite the item to correct the fault(s).

1. Alexander Graham Bell, who was born in Scotland in 1847 and moved to Ontario, is best known for his invention of the telephone, which he called his "electrical speech machine." T F
2. Minor differences among organisms of the same species are known as
  - A. heredity
  - B. variations
  - C. adaptation
  - D. natural selection
3. The recall of factual information can be best assessed with a \_\_\_\_\_ item.
  - A. essay
  - B. objective
  - C. matching
  - D. short-answer

4. Although the experimental research completed, particularly that by Hansmocker, must be considered too equivocal and the assumptions viewed as too restrictive, most testing experts would recommend that the easiest method of significantly improving written achievement test reliability would be to
  - A. increase the size of the group
  - B. increase the weighting of items
  - C. increase the number of items
  - D. increase the amount of testing time
  
5. An electric transformer can be used
  - A. alternating current is changed to direct current
  - B. for storing up electricity
  - C. it converts electrical energy into direct current
  - D. to increase the voltage of alternating current (correct answer)