PART II
Classroom Tests and Assessments

CHAPTER

6

PLANNING CLASSROOM TESTS AND ASSESSMENTS

Classroom tests and assessments play a central role in the evaluation of student learning. They provide relevant measures of many important learning outcomes and indirect evidence concerning others. They make expected learning outcomes explicit to students and parents and show what types of performances are valued. The validity of the information they provide, however, depends on the care that goes into the planning and preparation of tests and assessments.

The main goal of classroom testing and assessment is to obtain valid, reliable, and useful information concerning student achievement. This requires determining what is to be measured and then defining it precisely so that tasks can be constructed that require the intended knowledge, skills, and understanding while minimizing the influence of irrelevant or ancillary skills. It also requires specifying the achievement domain in such a manner that the sample of items and assessment tasks will represent the total domain of achievement tasks giving appropriate emphasis to high-priority objectives. Satisfying these requirements provides the foundation for obtaining results that will be valid for the intended instructional uses.

The likelihood of preparing valid, reliable, and useful classroom tests and assessments is greatly enhanced if a series of steps like that shown in Figure 6.1 is followed.

Specific procedures for constructing each of the various types of test items are described in Chapters 7 through 9. Open-ended essay questions are discussed in Chapter 10 and other types of performance assessment tasks used for measuring complex achievement are described in Chapter 11. The systematic collection of student work into portfolios and the use of those portfolios for specific purposes are described in Chapter 12. Chapter 13 describes the use of observational techniques, peer appraisals, and self-report

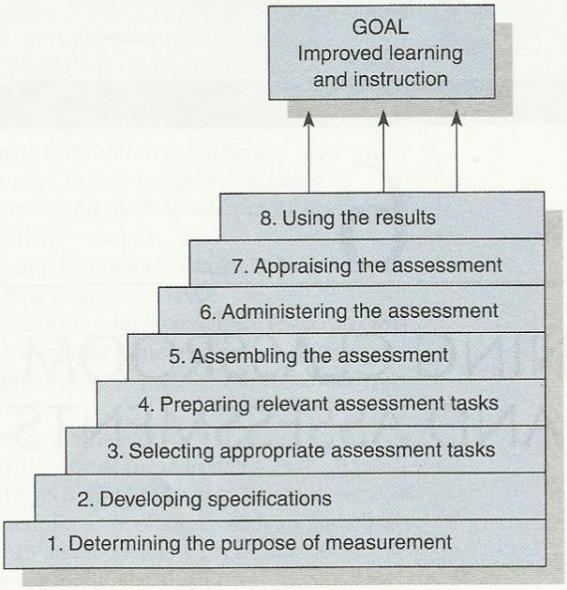

Figure 6.1
Basic steps in classroom testing and assessment

techniques to assess learning outcomes and student development. Chapter 14 is concerned with assembling, administering, and appraising classroom tests and assessments. Chapter 15, the last chapter in this section, describes approaches to grading and reporting.

THE PURPOSE OF CLASSROOM TESTING AND ASSESSMENT

Classroom tests and assessments can be used for a variety of instructional purposes. These can be best described in terms of their location in the instructional process, which closely parallels the types of assessment described in Chapter 3.

Pretesting

Tests and assessments may be given at the beginning of an instructional segment (e.g., unit or course) to determine (a) whether students have the prerequisite skills needed for the instruction (to determine readiness) or (b) to what extent students have already achieved the objectives of the planned instruction (to determine student placement or modification of instruction).

Readiness pretests are typically limited in scope. For example, a pretest in algebra might be confined to arithmetic operations and concepts, a pretest in physics might consist of

basic physical concepts or prerequisite skills in algebra, and a pretest in beginning German might be limited to knowledge of English grammar. In addition to being confined to a limited domain, pretest items tend to have a relatively low level of difficulty. Published readiness tests are available at the primary school level in the basic skill areas, but most teachers must develop their own readiness pretests. These tests serve as a basis for remedial work or for adaptation of instructional plans.

Pretests for determining the extent to which students have already achieved the objectives of the planned instruction are no different from the tests used to measure the outcomes of instruction. Thus, a test designed to measure final achievement in a course or unit may be given at the beginning to measure entry performance on the objectives. In this case, the final test or assessment should, of course, not be the same test used in pretesting but an equivalent form of it.

Testing and Assessment During Instruction

Tests and assessments given during instruction provide the basis for formative assessment. They are used to monitor learning progress, detect misconceptions, encourage students to study, and provide feedback to students and teachers. Teachers commonly call these formative tests learning tests, practice tests, quizzes, unit tests, and the like. These tests and assessments typically cover some predefined segment of instruction (e.g., a chapter or particular set of skills) and thus encompass a rather limited sample of learning outcomes. The mix of types of test items and more complex performance assessment tasks needs to be selected with care to ensure that the full range of critical instructional objectives is assessed and to ensure that high-priority but possibly difficult-to-assess objectives are adequately represented in the assessment. Ideally, the tests and assessments will be constructed in such a way that corrective prescriptions can be given for learning objectives that are yet to be achieved. When most students fail a set of items or perform poorly on an extended performance task, a group review may be applicable. When a small number of students perform in ways that show a lack of understanding of critical concepts, alternative methods of study may be prescribed (e.g., special tutoring, reading assignments, practice exercises, and the like).

Persistent learning difficulties may require the use of diagnostic tests. For this type of testing, a number of test items are needed in each specific area, with some slight variation from item to item. In diagnosing students' difficulties in adding whole numbers, for example, it would be necessary to include addition problems containing various number combinations, some requiring carrying and some not, to pinpoint the specific types of errors being made. Diagnostic testing is a highly specialized area that has been somewhat neglected in educational measurement. There are some published diagnostic tests, but these are primarily in the basic skills area. In other areas, teachers have to depend more heavily on the diagnostic features of formative tests or prepare their own diagnostic tests and assessments.

End-of-Instruction Testing and Assessment

At the end of a segment of instruction (e.g., unit or course), our main interest is in measuring the extent to which the intended learning outcomes and performance standards have been achieved. Although these end-of-instruction tests and assessments are used

Table 6.1
The basic types of classroom tests and assessments

	Timing				
	Before Instruction		During Instruction		End of Instruction
Function	*Readiness*	*Placement*	*Formative*	*Diagnostic*	*Summative*
Focus of measurement	Prerequisite entry skills	Course or unit objectives	Predefined segment of instruction	Most common learning errors	Course or unit objectives
Nature of sample	Limited sample of selected skills	Broad sample of all objectives	Limited sample of learning tasks	Limited sample of specific errors	Broad sample of all objectives
Item difficulty	Typically has low level of difficulty	Typically has wide range of difficulty	Varies with the segment of instruction	Typically has low level of difficulty	Typically has wide range of difficulty
Time of administration	Beginning of course or unit	Beginning of course or unit	Periodically during instruction	As needed during instruction	End of course or unit
Use of results	Remedy entry deficiencies or assignment to learning group	Instructional planning and advanced placement	Improve and direct learning through ongoing feedback	Remedy errors related to persistent learning difficulties	Assign grades, certify accomplishment, or evaluate teaching

Source: Adapted from "Functional Types of Student Evaluation" by P. W. Airasian and G. F. Madaus, 1972, *Measurement and Evaluation in Guidance*, 4, pp. 221–233.

primarily for summative assessment (e.g., to certify accomplishment or assign grades), they can also serve other functions. End-of-unit tests can be used for feedback to students, encouraging students to undertake more challenging advanced work, assigning of remedial work, and assessing instruction as well as for grading purposes. In fact, they can serve the functions of both formative and summative assessment and in some cases serve as a pretest for the following unit (e.g., where the units are sequenced, as in math). End-of-course tests provide a broad survey of student learning over all the intended outcomes of a course. In addition to their use in grading, tests can also provide information for evaluating instructional effectiveness. A summary of the various types of classroom tests is presented in Table 6.1.

DEVELOPING SPECIFICATIONS FOR TESTS AND ASSESSMENTS

To ensure that classroom tests and assessments measure a representative sample of instructionally relevant tasks, it is important to develop specifications that can guide the selection of test items and assessment tasks. When preparing a brief learning test on a

limited area (e.g., spelling or capitalization), a simple listing of specific tasks, with an indication of the number of items measuring each task, may suffice. For most testing purposes, however, more elaborate specifications are needed. One device that has been widely used for this purpose is a two-way chart, called a table of specifications (or test blueprint).

Building a Table of Specifications

Building a table of specifications involves (a) preparing a list of instructional objectives, (b) outlining the course content, and (c) preparing the two-way chart. These steps are illustrated here for a weather unit in a middle school earth science class.

Preparing the List of Instructional Objectives. Using the procedure for stating instructional objectives described in Chapter 3, the following list of general objectives and specific learning outcomes was prepared for the weather unit. Although the list is not exhaustive, it illustrates both the method of stating the objectives and the desired amount of detail. If this weather unit were being taught at a lower grade level, our objectives would be defined by fewer and simpler specific learning outcomes; at a higher grade level, more complex learning outcomes would be included.

Instructional Objectives for a Weather Unit

1. Knows basic terms
- 1.1 Writes a definition of each term
- 1.2. Identifies the term that represents each weather element
- 1.3. Identifies the term that best fits a given weather description
- 1.4. Matches the term to a picture of the concept
- 1.5. Distinguishes between correct and incorrect uses of the term

2. Knows weather symbols
- 2.1. Matches each symbol with the weather element it represents
- 2.2. Draws the symbol for each weather element
- 2.3. Identifies the meaning of each symbol

3. Knows specific facts
- 3.1. Identifies the elements affecting weather
- 3.2. Names the instrument used for measuring each weather element
- 3.3. Identifies the unit of measurement used in reporting each weather element
- 3.4. Distinguishes between correct and incorrect procedures for determining each weather element
- 3.5. Matches the names of each cloud type with a description of its characteristics
- 3.6. Identifies the weather conditions associated with each type of front

4. Understands the influence of factors on weather formation
- 4.1. Describes the characteristics of a given weather condition
- 4.2. Describes how clouds are formed
- 4.3. Distinguishes between probable and improbable weather reports
- 4.4. Identifies the factors causing a given weather change
- 4.5. Predicts future weather from a change in conditions
- 4.6. Explains how clouds affect weather and climate
- 4.7. Delivers an oral weather report for another state to class

5. Interprets weather maps
   - 5.1. Describes the weather for a given locality
   - 5.2. Identifies the different types of fronts shown on a weather map
   - 5.3. Describes the weather conditions surrounding each front shown on a weather map
   - 5.4. Identifies the direction of movement for each front shown on a weather map
   - 5.5. Explains ways global patterns of atmospheric movement influence local weather
   - 5.6. Constructs a weather map using a written description of existing conditions
6. Measures relevant weather characteristics
   - 6.1. Measures and records changes in air pressure
   - 6.2. Measures and records direction and velocity of wind
   - 6.3. Measures relative humidity

The list of objectives is limited to those outcomes that can be measured by a classroom assessment. The form of the test question or assessment task may vary as a function of the outcome. Outcomes involving verbs such as *identifies*, *distinguishes*, or *matches* are effectively and efficiently measured by multiple-choice or other types of objective test items. Those involving verbs such as *names* or *draws* are readily measured by simple supply-type items where students are asked to "fill in the blank." Outcomes using verbs such as *describes* or *explains* can be measured by skillfully constructed objective items but are often more readily and better measured by essay questions. Those involving verbs such as *constructs*, *delivers*, or *measures* most clearly require performance assessment tasks. The listed objectives do not include affective outcomes (e.g., demonstrating scientific attitude). A more elaborate evaluation plan for a weather unit might contain a listing of all relevant outcomes with an indication of how each is to be assessed.

Outlining the Instructional Content. The second step in preparing a table of specifications is to make an outline of the instructional content. The instructional objectives describe the types of performance the students are expected to demonstrate; the instructional content indicates the area in which each type of performance is to be shown. The amount of detail to include in the content outline is somewhat arbitrary, but the outline should be detailed enough to ensure an adequate sampling of content and proper interpretation of the results. The following list of topics and subtopics for a weather unit illustrates an adequate outline for test and assessment preparation.

Content Outline for a Weather Unit

A. Air pressure
   1. Measuring and reporting air pressure
   2. Factors affecting air pressure
   3. Relation to weather changes
B. Air temperature
   1. Measuring and reporting air temperature
   2. Factors affecting air temperature
   3. Relation to weather formations
C. Humidity and precipitation
   1. Measuring and reporting humidity
   2. Factors affecting humidity

3. Forms of precipitation
4. Measuring and reporting precipitation

D. Wind
1. Measurement of speed and direction
2. Factors affecting speed and direction
3. Symbols for reporting speed and direction

E. Clouds
1. Types of clouds
2. Characteristics of cloud types
3. Factors causing cloud formations
4. Relation to weather conditions
5. Symbols for cloud types

F. Fronts
1. Types of fronts
2. Formation of fronts
3. Weather related to fronts
4. Symbols for fronts

Preparing the Two-Way Chart. The final step in building a table of specifications is to prepare a two-way chart that relates the instructional objectives to the instructional content, thus specifying the nature of the sample of test items and assessment tasks. An example of the chart for our junior high weather unit is presented in Table 6.2. This table indicates both the total number of test items and assessment tasks and the percentage allotted to each objective and each area of content. For classroom testing, using the number of items may be sufficient, but the percentages are useful in determining the amount of emphasis to give to each area.

Table 6.2
Table of specifications for a weather unit in a middle school earth science class

	Objectives						
	---------Knows---------			Understands	Interprets		
Content	*Basic Terms*	*Weather Symbols*	*Specific Facts*	*Influence of Each Factor on Weather Formation*	*Weather Maps*	*Total Number of Items*	*Percentage of Items*
Air pressure	1	1	1	3	3	9	15
Wind	1	1	1	10	2	15	25
Temperature	1	1	1	4	2	9	15
Humidity and precipitation	1	1	1	7	5	15	25
Clouds	2	2	2	6		12	20
Total number of items	6	6	6	30	12	60	
Percentage of items	10	10	10	50	20		100

Table 6.2 is limited to those objectives that can be measured with a classroom test with objective or restricted-response essay items. It excludes those objectives requiring performance (e.g., 4.7, an oral presentation; 5.6, constructing a weather map; and 6.1 through 6.3, using weather instruments). Thus, for comprehensive coverage, the classroom test would need to be supplemented by a set of performance assessment tasks. The relative weight to be given to the results of the performance assessments and the classroom test in the assignment of course grades should be determined in accordance with the relative importance of the objectives. It is critical that important objectives that cannot be assessed by the classroom test be identified so that plans for assessing the achievement of those objectives can be made. In this example, a simple list of five performance assessment tasks corresponding to instructional objectives 4.7, 5.6, 6.1, 6.2, and 6.3, together with a plan for the relative weight to be given to the scores on those tasks and the classroom test, would suffice.

Table 6.2 was prepared using the following steps.

1. List the general instructional objectives across the top of the table.
2. List the major content areas down the left side of the table.
3. Determine what proportion of the test items should be devoted to each objective and each content area.

The bottom row of Table 6.2 shows that 10% of the items are to be devoted to knowledge of basic terms, 10% to knowledge of weather symbols, 10% to knowledge of specific facts, 50% to understanding influence of weather factors, and 20% to interpretation of weather maps. Similarly, the column on the far right shows that 15% of the items are to be concerned with air pressure, 25% with wind, and so on down the column. Typically, these percentages are assigned first to indicate the relative emphasis to be given to each objective and each content area. The total number of items for each area is then determined. For example, because this is to be a 60-item test, 6 of the items (10%) are to measure knowledge of basic terms, and 9 of the items (15%) are concerned with air pressure. Thus, the total number of items is computed and listed across the bottom and down the right-hand side of the chart. These numbers are then used as a guide to assign test items to each cell. For example, 2 items (approximately 33% of 6) have been assigned to the cell in the lower-left-hand corner, indicating that the test will contain 2 items measuring knowledge of terms concerning clouds. The assignment of items to each cell is not strictly a matter of following the percentages, however. Some cells may be left blank, as shown in the chain, because items in those areas are inappropriate. Similarly, some cells may receive a larger number of items because they correspond to more important instructional objectives, and consequently the learning tasks for that cell may have received greater emphasis in teaching. In the example, 10 of the 60 items would be allocated to assessing a student's understanding of wind on weather formation, whereas only a single item would be included to measure knowledge of a specific fact about wind. Such a disproportionate allocation is appropriate if it is consistent with the intended and actual instructional priority given to understanding the influences of wind on weather formation. The table of specifications provides guidance in constructing the test. The actual test may deviate slightly from those specifications for specific cells, but the number of items in the cells of the table should approximate the desired distribution and should add up to the total number of items indicated for each column and row.

The final distribution of items in the table of specifications combined with performance assessments that are not part of the classroom tests should reflect the emphasis given during the instruction. Objectives considered more important by the teacher should be allotted more test items. This applies not only to the items within the classroom test but also to performance assessment tasks. The weight given to the performance of such assessment tasks should reflect the importance of the objective. Similarly, areas of content receiving more instruction time should be allocated more test items and assessment tasks. Although the decisions involved in making the table are somewhat arbitrary and the process is time consuming, the preparation of a table of test specifications, together with a listing of performance assessment tasks needed to supplement the test, is one of the best means for ensuring that the total set of test items and assessment tasks will measure a representative sample of instructionally relevant objectives. As noted in Chapter 4, this provides content-related evidence that the test will provide a valid measure of the intended learning outcomes. The effort that goes into developing specifications also makes it much easier to prepare the test once the plan is developed.

Other Examples of Specifications

Because of its broad coverage, a table of specifications is especially useful in constructing a test or assessment over a unit or course. Here the table ensures that each of the diverse types of learning outcomes will receive appropriate emphasis in the test. Whether a table of specifications is useful in constructing a test over a more limited area, however, depends on the scope of learning outcomes to be covered by the test. If the domain of tasks is very limited, such as "addition of fractions with the same denominator," a table of specifications might be unnecessary. Here we could simply list all, or nearly all, of the specific tasks encompassed by this learning outcome, as follows:

- Adds two fractions with same denominator where the answer is less than 1 (1/3 + 1/3).
- Adds two fractions with same denominator where the answer equals 1 (1/3 + 2/3).
- Adds two fractions with same denominator where the answer is greater than 1 (2/3 + 2/3).
- Adds two fractions with same denominator and reduces answer to lowest terms (4/6 + 4/6).
- Adds more than two fractions with same denominator and reduces answer (1/8 + 5/8 + 6/8).

A list of learning tasks such as this specifies quite clearly the precise nature of the performance involved in "the addition of fractions with the same denominator." We could obtain a fairly representative sample of such tasks by simply constructing five or more items for each task, using various number combinations.

If our mathematics test were to cover a slightly larger achievement domain, say, "addition of fractions," we might now find a table of specifications quite useful. An illustration of such a table for a 50-item test is shown in Table 6.3. The use of such a table does not mean, of course, that we should not make the type of detailed breakdown we did for

Table 6.3
Table of specifications for a 50-item test on addition of fractions

	Instructional Objectives			
Content Area	*Adds Fractions*	*Adds Fractions and Mixed Numbers*	*Adds Mixed Numbers*	*Total Items*
Denominators are alike	5	5	5	15
Denominators are unlike (with common factor)	5	5	5	15
Denominators are unlike (without common factor)	6	7	7	20
Total items	16	17	17	50

"addition of fractions with the same denominator." Such a detailed listing of tasks aids in teaching, in detecting learning errors, and in constructing test items. The table of specifications, however, supplements such lists by specifying the sample of tasks to be included in the test. As noted earlier, this provides greater assurance that the intended learning outcomes will be measured in a balanced manner.

A simple list such as that given for "addition of fractions with the same denominator" or specifications such as those in Table 6.3 can be readily constructed for the operational skills aspects of addition of fractions. It is often more difficult, however, to construct specifications that do justice to objectives of understanding concepts or the ability to apply the concepts in novel settings. This is made clear even with the relatively narrow topic of fractions by considering the National Council of Teachers of Mathematics (NCTM) Standard 12: Fractions and Decimals for Grades K–4. Among other things, that standard states that students should "develop concepts of fractions, mixed numbers, and decimals" and "apply fractions and decimals to problem situations" (NCTM, 1989).

Using Standard 12 as a statement of instructional goals for fractions, the assessment would clearly need to include tasks that measure students' conceptual understanding of fractions and their ability to apply fractions in novel problem situations. To ensure the inclusion of such tasks, understanding concepts and applications of fractions should be included in the table of specifications. Although the specifications in Table 6.3 might be appropriate for one segment of instruction, specifications similar to those in Table 6.4 might be more appropriate at another stage of instruction.

Because Table 6.4 is expressed in terms of the percentage of points associated with each combination of objectives and content area, a decision would need to be made about the total number of points. Because the points are to be obtained from a combination of a classroom test and a set of performance assessment tasks, additional specifications would also be needed to allocate the points to the different sources of information. A single performance assessment task involving relationships between decimals and fractions, for example, might be scored on a 5-point scale. Assuming a total

Table 6.4
Table of specifications for fractions and decimals showing targeted percentage of points from a combination of a classroom test and a set of performance assessment tasks

	Instructional Objectives			
Content Area	*Procedural Skills*	*Understanding*	*Application*	*Total Percentage of Points*
Simple fractions	5	10	5	20
Mixed numbers	5	15	10	30
Decimals	5	10	5	20
Decimal–fraction relationships	10	10	10	30
Total percentage of points	25	45	30	100

Table 6.5
List of reading comprehension skills and number of items for each specific skill

Reading Skill	*Number of Items*
Identifies details stated in a passage	10
Identifies the main idea of a passage	10
Identifies the sequence of actions or events	10
Identifies relationships expressed in a passage	10
Total number	50

of 100 points, that task would represent half the targeted total of 10 points for that cell. The other 5 points might then be allocated to five short-answer test items.

Using a One-Way Classification System

For tests in some areas, a one-way classification of items may be all that is needed. In planning for a reading test, for example, a list of the reading skills and the number of test items for measuring each skill may be sufficient for specifying what the test is to measure. The content (e.g., passages read) may vary from time to time, but the skill outcomes remain fairly constant. Thus, a master list of skills can be prepared for use with various types of reading material. An example of such a list is presented in Table 6.5.

It should be noted that each skill in Table 6.5 is stated in specific performance terms and that 10 items are used to measure each skill. This provides for criterion-referenced interpretation. Although the material to be read is not included in the specifications, it will, of course, need to be carefully selected in terms of interest and readability level. See the "Checklist" box.

CHECKLIST

Reviewing Specifications for Tests and Assessments

	Yes	No
1. Are the specifications in harmony with the purpose of the test or assessment?	____	____
2. Do the specifications indicate the nature and limits of the achievement domain?	____	____
3. Do the specifications indicate the types of learning outcomes to be measured?	____	____
4. Do the specifications indicate the sample of learning outcomes to be measured?	____	____
5. Is the number of test items of assessment tasks indicated for each subdivision?	____	____
6. Are the types of items and tasks to be used appropriate for the outcomes to be measured?	____	____
7. Is the distribution of items and tasks adequate for the types of interpretation to be made?	____	____
8. If sample items or tasks are included, do they illustrate the desired attributes?	____	____
9. Do the specifications, as a whole, indicate a representative sample of instructionally relevant tasks that fits the use to be made of the results?	____	____

SELECTING APPROPRIATE TYPES OF ITEMS AND ASSESSMENT TASKS

It is common to make a distinction between classroom tests that consist of objective test items and performance assessments that require students to construct responses (e.g., write an essay) or perform a particular task (e.g., measure air pressure). Objective test items are highly structured and require the students to supply a word or two or to select the correct answer from alternatives. They are called objective because they have a single right or best answer that can be determined in advance. Performance assessment tasks, such as essay questions, permit the student to organize and construct the answer in essay form. Other types of performance assessment tasks may require the student to use equipment, generate hypotheses, make observations, construct something (e.g., a model), or perform for an audience (e.g., give a speech). For most performance assessment tasks, there is not a single right or best response—there may be a variety of responses that are considered excellent. Problem identification may, in fact, be an important part of the task, and, of course, there may be multiple ways of structuring the problem and organizing a response. Expert judgment is required to score the performances. There is no

conflict between these highly constrained objective items and the much less constrained performance assessment tasks. For some instructional purposes, objective items may be most efficient, whereas for others, performance assessments may prove most satisfactory. Each approach should be used where most appropriate, with appropriateness determined by the learning outcomes to be measured and by the advantages and limitations of each approach.

The Objective Test Item

Various forms and uses of objective items are discussed in greater detail in Chapters 7 to 9. Here we provide only a quick overview.

The great variety of different types of objective test items can be classified into those that require the student to supply the answer and those that require the student to select the answer from a given number of alternatives. These two general classes are commonly divided into basic types of objective test items, illustrated in the accompanying examples.

EXAMPLES

SUPPLY TYPES

Short Answer

What is the name of the author of *Moby Dick? (Herman Melville)*
What is the formula for hydrochloric acid? (*HCl*)
What is the value of X in the equation $2X + 5 = 9$? (*2*)

Completion

Lines on a weather map joining points with the same barometric pressure are called_____. (*isobars*)

The formula for ordinary table salt is_____. (*NaCl*)
In the equation $2X + 5 = 9$, $X =$ _____(*2*)

SELECTION TYPES

Matching

(C)	1. And	A. Adjective
(D)	2. Dog	B. Adverb
(G)	3. Jump	C. Conjunction
(F)	4. She	D. Noun
(B)	5. Quickly	E. Preposition
		F. Pronoun
		G. Verb

True–False or Alternative Response

(T)	F	A virus is the smallest known organism.
T	(F)	An atom is the smallest particle of matter.
Yes	(No)	In the equation $2X + 5 = 9$; X equals 3.
(Yes)	No	Acid turns litmus paper red.

Multiple Choice

In the equation $2X + 5 = 9$, $2X$ means

A 2 plus X.
B 2 minus X.
C 2 divided by X.
(D) 2 multiplied by X.

In which of the following sentences do the subject and verb disagree?

A When they win, they are happy.
(B) Politics are hard to understand.
C The majority is always right.
D One or the other is to be elected.

In addition to these basic types of objective test items are numerous modifications and combinations of types. However, little is to be gained from listing all the possible variations, as many are unique to particular objectives or subject-matter areas. Some of the more common variations used to measure understanding, thinking skills, and other complex learning outcomes are illustrated later. An understanding of the principles of test construction and the principles that apply to each of the specific types of objective test items should enable teachers to make adaptations that best fit their particular purposes.

The various types of objective test items have one feature in common that distinguishes them from performance assessment tasks: They present students with a highly structured task that limits the type of response they can make. To obtain the correct answer, students must demonstrate the specific knowledge, understanding, or skill called for in the item; they are not free to redefine the problem or to organize and present the answer in their own words. They must select one of several alternative answers or supply the correct word, number, or symbol. This structuring of the problem and restriction on the method of responding contribute to objective scoring that is quick, easy, and accurate. On the negative side, this same structuring makes the objective test item inappropriate for measuring the ability to formulate problems and choose an approach to solving them or the ability to select, organize, and integrate ideas. To measure such outcomes, we must depend on performance assessment tasks. Performance assessment tasks are also needed to measure a student's ability to engage in hands-on activities, such as conducting an experiment, measuring precipitation, designing and conducting a survey, or creating an art object.

Because written essays are the most commonly used form of performance assessment tasks, it is worth focusing on essay questions before considering other types of performance assessment tasks. The essay question is commonly viewed as one item type. A useful classification, however, is one based on the amount of freedom of response the student is allowed. This includes extended-response essay questions, in which students are given almost complete freedom in making their responses, and restricted-response essay questions, in which the nature, length, or organization of the response is limited. These types of essay questions are illustrated as follows:

EXAMPLES **EXTENDED-RESPONSE ESSAY QUESTIONS**

Describe what you think the role of the federal government should be in maintaining a stable economy in the United States. Include specific policies and programs and give reasons for your proposals.

Explain how clouds are formed and how they affect weather and climate.

Predict the effects that an increase in interest rates would have on consumers, producers, and investors. Include an explanation of how your predictions follow from fundamental economic principles.

RESTRICTED-RESPONSE ESSAY QUESTIONS

State two advantages and two disadvantages of maintaining high tariffs on goods from other countries.

Name two types of clouds and describe the characteristics that distinguish them.

The percentage of the workforce that is unemployed increased following an increase in interest rates. Briefly explain the relationship.

Extended-response essay questions permit students to decide which facts they think are most pertinent, to select their own method of organization, and to write as much as seems necessary for a comprehensive answer. Thus, such questions tend to reveal the ability to evaluate ideas, to relate them coherently, and to express them succinctly. To a lesser extent, they also reflect individual differences in attitudes, values, and creative ability.

Although extended essay questions are valuable for measuring complex skills and understanding of concepts and principles, they have three weaknesses: (1) They are inefficient for measuring knowledge of factual material because the questions are so extensive that only a small sample of content can be included in any one assessment; (2) the scoring criteria are not as apparent to the student because it is not listed with the item; and (3) scoring is difficult and apt to be unreliable because the responses include an array of factual information of varying degrees of correctness, organized with varying degrees of coherence, and expressed with varying degrees of legibility and conciseness. Hence, extended essay questions are best used to assess more complex cognitive objectives involving application, evaluation, or the ability to express ideas in coherent, persuasive ways.

Restricted-response essay questions minimize some of the weaknesses of extended-response essay questions. Restricting the type of response makes it easier to measure

knowledge of factual material, makes the scoring more clear to the student, and reduces somewhat the difficulty of the scoring. On the other hand, the more highly structured task presented by the restricted-response essay questions makes them less effective as a measure of the ability to select, organize, and integrate ideas. Furthermore, if the restrictions become too tight, the questions reduce to nothing more than a supply-type item.

Neither extended-response essay questions nor restricted-response essay questions can serve all purposes equally. The type of essay question to use in a particular situation depends primarily on the learning outcomes to be measured and, to a lesser extent, on such practical considerations as the difficulty of scoring.

Other Types of Performance Assessment

There are many other types of performance assessment tasks. Examples include oral presentations; construction of graphs, diagrams, or models; use of equipment or scientific instruments; typing; and playing a musical instrument. Tasks are selected so that the performance corresponds as closely as possible to an important instructional objective.

A distinction can be made between restricted-response performance tasks and extended-response performance tasks. These types are illustrated as follows:

EXAMPLES

EXTENDED-RESPONSE PERFORMANCE TASKS

Prepare a weather report for another state and make an oral presentation of the report to the class using appropriate visual displays.

Take four strips of different-colored paper that are each 6 inches in length. Cut the white strip into two parts of equal length, the yellow strip into three parts of equal length, the green strip into four parts of equal length, and the blue strip into six parts of equal length. Compare the lengths when different numbers of pieces of strips are put together (e.g., two pieces of the yellow strip compared to three pieces of the blue strip) and describe the relationship between corresponding fractions (e.g., 2/3 and 3/6).

RESTRICTED-RESPONSE PERFORMANCE TASKS

Measure and record the relative humidity.

Take a number of 1-inch-square pieces of paper to use. Assuming each square represents the fraction 1/4, construct a whole. (Can be repeated using different fractions.)

The virtues and limitations of extended- and restricted-response performance tasks are much the same as those listed for extended- and restricted-response essay questions. The freedom provided by the extended-response performance task enables students to display such important skills as problem solving, planning, organization, integration, and creativity. Questions of this type also provide an opportunity to observe student performance in more realistic contexts than are possible with objective test items. On the other hand, they are time consuming to administer and difficult to score.

Restricted-response performance tasks are generally easier to score and require less time than extended-response performance tasks. However, they are generally less suited

for measuring the higher-order skills measured by extended-response performance tasks.

Comparative Advantages of Objective Test Items and Performance Assessment Tasks

Both objective test items and performance assessment tasks can provide valuable evidence concerning student achievement. Each has advantages and limitations that make it more appropriate for some purposes than for others. A comparison of the relative merits of objective item tests and performance assessments is presented in Table 6.6. In considering the comparative advantages of objective-item tests and performance assessments, we must be careful not to fall into either/or thinking, that is, to use either objective items or performance assessment tasks. It is frequently better to use both, with each measuring the particular learning outcomes for which it is best suited. This should also have a desirable influence on student learning because in preparing for such tests students must attend to the specific types of learning outcomes measured by objective items as well as the synthesis and integrated performance outcomes measured by performance assessment tasks.

Selecting the Most Appropriate Types of Items and Tasks

A basic principle in selecting the type of test item and assessment task to use is to **select the item type that provides the most direct measure of the intended learning outcome**. If, for example, the intended learning outcome is writing, naming, listing, or speaking, the task should require the students to supply the answer. If the outcome involves the use of laboratory equipment to solve a problem, nothing short of an actual laboratory performance task will suffice. On the other hand, if the task calls for idntifying a correct answer, a selection-type item should be used. In those cases where the specific learning outcome does not make clear which item type to use, selection-type items would be favored because of the greater control over the student's response and the objectivity of the scoring. However, the intended learning outcome may also require more complex items or direct measures of performance. The following items were prepared for the weather unit discussed earlier. Note how each type of objective item provides a direct measure of the outcome it was designed to measure.

EXAMPLE **SHORT-ANSWER ITEMS**

Specific Learning Outcome: Writes a definition of each term.

Directions: Write a one-sentence definition of each of the following terms.

1. Weather
2. Humidity
3. Occluded front

Specific Learning Outcome: Names the instrument used for measuring each weather element.

1. The instrument used to measure the amount of precipitation in a given locality is called a (an) _____.

Table 6.6
Comparative advantages of objective tests and performance assessments

	Objective Test	*Performance Assessment*
Learning outcomes measured	Is efficient for measuring knowledge of facts. Some types (e.g., multiple choice) can also measure understanding, thinking skills, and other complex outcomes.	Can measure understanding, thinking skills, and other complex learning outcomes (especially useful where originality of response is desired).
	Inefficient or inappropriate for measuring ability to select and organize ideas, writing abilities, and some types of problem-solving skills.	Appropriate for measuring performance on tasks corresponding to important instructional objectives in realistic contexts. Is inefficient for measuring knowledge of facts.
Preparation of questions	A relatively large number of questions are needed for a test.	Only a few tasks are needed for an assessment.
	Preparation is difficult and time consuming.	
Sampling of course content	Provides an extensive sampling of coursecontent because of the large number of questions that can be included in a test.	Sampling of course content is usually limited because of the small number of tasks that can be included in an assessment.
Control of student response	Complete structuring of task limits students to type of response called for. Prevents bluffing and avoids influence of writing skill, though selection-type items are subject to guessing.	Freedom to respond in own way enables students to display originality, and guessing is minimized.
Scoring	Objective scoring.	Judgmental scoring.
Influence on learning	Usually encourages student to develop a comprehensive knowledge of specific facts and the ability to make fine discriminations among them. Can encourage the development of understanding, thinking skills, and other complex outcomes if properly constructed.	Encourages students to concentrate on larger units of subject matter, with special emphasis on the ability to organize, integrate, and express ideas effectively.
Reliability	High reliability is possible and is typically obtained with well-constructed tests.	Reliability is typically low, primarilybecause of a limited sample of tasks and inconsistent scoring.

Specific Learning Outcome: Lists the characteristics of a given weather phenomenon.

1. List three main characteristics of a hurricane.

Specific Learning Outcome: Measures relative humidity.

1. Use the appropriate instrument to measure the relative humidity and record the value obtained.

MULTIPLE-CHOICE ITEMS

Specific Learning Outcome: Identifies the units of measurement used in reporting each weather element on a weather map.

1. United States weather maps indicate air pressure in

A inches
B feet
(C) pounds
D millibars

TRUE–FALSE ITEMS

Specific Learning Outcome: Distinguishes between correct and incorrect procedures for determining each weather element.

T	(F)	1.	Dew point is determined by cooling a sample of air until it is free of moisture.
(T)	F	2.	Ceiling is determined by using balloons that rise at known rates.

MATCHING ITEMS

Specific Learning Outcome: Matches each weather instrument to the weather element it measures.

Directions: On the line to the left of each weather element in Column A, write the letter of the weather instrument in Column B that is used for measuring it. Each instrument in Column B may be used once, more than once, or not at all.

Column A	*Column B*
(B) 1. Air pressure	A. Anemometer
(E) 2. Air temperature	B. Barometer
(C) 3. Humidity	C. Hygrometer
(A) 4. Wind velocity	D. Rain gauge
	E. Thermometer
	F. Wind vane

PERFORMANCE TASKS

Specific Learning Outcome: Measures and records changes in air pressure.

Directions: Measure the air pressure on 5 consecutive days at the same time of day and graph it to see the changes. Explain the reasons for the changes that occur.

In deciding which selection-type item to use, a common practice is to use the multiple-choice item if it will measure the learning outcome as directly as the other two types. The use of true–false items is typically most valuable in those special instances where there are only two possible alternatives (e.g., distinguishing between correct and incorrect procedures). The matching item is a specialized form of the multiple-choice item and should be used only where a series of homogeneous things are to be related (e.g., dates and events, authors and books, or instruments and uses). The multiple-choice item is favored for most other selection-type tasks because the use of four or five alternatives reduces the chances of guessing the answer and provides clues to students' misunderstandings. See Appendix G for other examples of how to relate test items to intended learning outcomes.

Whether a test item or an assessment task actually measures the particular performance called for by a specific learning outcome depends, of course, to a large extent on the skill with which the test item or assessment task is constructed. No amount of skill, however, will enable us to develop a valid test or assessment of achievement if the items or tasks selected for use are inappropriate for measuring the intended outcomes.

CONSIDERATIONS IN PREPARING RELEVANT TEST ITEMS AND ASSESSMENT TASKS

The construction of items and tasks should be preceded by a series of preliminary steps. First, the purpose of the test or assessment should be determined. Second, a set of specifications should be developed. Third, the most appropriate types of test items and performance assessment tasks should be selected. Finally, items and tasks should be constructed in accordance with the specifications developed during the preceding steps. The rules for constructing each type of objective item are discussed in Chapters 7 through 9, and rules for constructing performance assessments are discussed in Chapters 10 and 11. Here we focus on general considerations involved in preparing relevant objective items and assessment tasks.

Matching Items and Tasks to Intended Outcomes

Classroom tests and assessments are most likely to provide a valid measure of the instructional objectives if the test items and assessment tasks are designed to measure the

performance defined by the specific learning outcomes. The process of matching test items and assessment tasks to the learning outcomes to be measured was illustrated earlier. Essentially, it involves fitting each item or task as closely as possible to the intended outcome, as follows:

EXAMPLE *Specific Learning Outcome:* Identifies the function of a given body structure.

Relevant Test Item:

What is the function of the kidneys?

(A) Eliminate waste products
B Improve the circulation of blood
C Maintain respiration
D Stimulate digestion

Thus, the preparation of relevant test items and assessment tasks means analyzing the performance described in the specific learning outcome (i.e., "Identifies the function of . . . ") and constructing a test item or assessment task that calls forth that performance (i.e., "What is the function of . . . ?"). Note that the specific learning outcome defines the type of response the student is expected to make, but it does not indicate the specific course content (i.e., kidney) the student is to identify. Keeping the learning outcome free of specific course content makes it possible to key the intended response to various areas of content. For example, students could be asked to identify the function of the heart, the lungs, the muscles, or any other body structure pertinent to the course's content. The desired student performance stated in the specific learning outcome can be keyed to each specific area of content by means of the table of specifications.

In some cases, it may be desirable to prepare a general item pattern as an intermediate step between the specific learning outcome and the test item. A general item pattern for our illustrative test item would be as follows:

EXAMPLE What is the function of . . . ?

An item pattern such as this could be completed adding the name of any body structure and using it as a short-answer question or by listing appropriate alternatives and using it as a multiple-choice item. Thus, using the item pattern as a guide, we could generate large numbers of relevant test items for this particular learning outcome. This procedure is especially useful when a file of test items is being prepared or when more than one form of the test is needed (e.g., pretesting, posttesting, and retesting).

When item patterns are used as a guide to test construction, they can be arranged by general type of learning outcome.

EXAMPLES **Knowledge Outcomes**

1. What is the name of . . . ?
2. What is the location of . . . ?
3. What are the characteristics of . . . ?
4. What is the function of . . . ?

Understanding Outcomes

1. What is the reason for . . . ?
2. What is the relationship between . . . ?
3. Which of these is an example of . . . ?
4. Which of these best summarizes . . . ?

Application Outcomes

1. What method would be best for . . . ?
2. What steps should be followed to construct . . . ?
3. Which of these indicates correct application of . . . ?
4. Which of these solutions is correct for . . . ?

Item patterns such as these should not, of course, be developed haphazardly. Rather, they should be derived from the specific learning outcomes they represent. Although it usually will not be possible to develop item patterns for all outcomes, listing them will help generate pools of relevant test items. The test construction time saved by using such a list can then be used to construct more effective items for those areas in which general item patterns are not feasible.

Obtaining a Representative Sample of Items and Tasks

A test or assessment, no matter how extensive, is almost always a sample of the many possible test items or tasks that could be included. We expect students to know thousands of facts, but we can test for only a limited number of them; we expect students to develop understanding applicable to innumerable situations, but we can test application to only a limited number of situations; and we expect students to develop thinking skills that will enable them to solve a variety of problems, but we can test their problem-solving ability with only a limited number of problems. In each area of content and for each specific learning outcome, then, we merely select a sample of student performance and accept it as evidence of achievement in that area. We assume that the students' responses to our selected set of items and tasks are typical of what their responses would be to other items and tasks drawn from the same area. This means, of course, that our limited samples must be selected in such a way that they provide as representative a sample as possible in each of the various areas for which the test or assessment is being developed.

Our sampling is most likely to be representative when the preparation of a test or assessment is guided by a carefully prepared set of specifications. Unless a table of specifications or some similar device is used as a guide in construction, there is a tendency

to overload the test with items measuring knowledge of isolated facts and to neglect more complex learning outcomes. In social studies, for example, it is not uncommon to include a disproportionately large number of items that measure knowledge of names, dates, places, and the like. In science, defining terms and naming structures and functions are commonly overemphasized. In mathematics, computational skill is frequently the only learning outcome measured. In language arts and literature, the identification of parts of speech, literary characters, authors, and the like is frequently too prominent. These learning outcomes are stressed not because we think that knowledge of isolated facts is more important than understanding, applications, interpretations, and various thinking skills. Rather, they usually receive undue prominence because we find it easier to construct this kind of test item. Without a carefully developed test plan, ease of construction all too frequently becomes the dominant criterion in constructing test items. As a consequence, the test measures a limited and biased sample of learning tasks and neglects many learning outcomes of greater importance.

Supplementing objective test questions with performance assessment tasks is one means of ensuring that broader learning objectives are given proper attention. However, it is inefficient to rely only on performance assessments to measure more complex understandings. For this reason, it is important to find ways of constructing objective test items that do more than measure factual knowledge. One approach to this goal of measuring complex achievement is through the use of interpretive exercises. This approach is discussed in detail in Chapter 9.

Number of Items and Tasks. The number of items and tasks is, of course, an important factor in obtaining a representative sample. The number of items and the number of performance tasks are determined when the set of specifications is built and depend on such factors as the purpose of measurement, the types of test items and assessment tasks used, the age of the students, and the level of reliability needed for effective use of the test or assessment results. Thus, an assessment over a third-grade social studies unit might contain 30 objective items, whereas a survey test over a 10th-grade social studies course might contain more than 100 objective items and several essay questions. Although there are no hard-and-fast rules for determining the number of items and tasks, an important consideration from a sampling standpoint is the number of items or tasks devoted to each specific area being measured. We want our tests and assessments to be long enough to provide an adequate sampling of each objective and each content area. As a rule of thumb, it is desirable when constructing a unit test to use at least 10 objective test items to measure each specific learning outcome. This number, however, might be lowered to as few as five if the task is extremely limited (e.g., "Adds two single-digit numbers" or "Capitalizes proper names") and the students are to supply the answers rather than to select them. For a survey test, where the sample of test items typically covers a broad area and emphasis is on the total score, using several objective test items for each specific learning outcome and 10 or more for each general objective probably would be sufficient.

Special problems of sampling arise when complex learning outcomes are being measured, because this requires more elaborate objective-type items (referred to as interpretive exercises) and performance assessment tasks. Both interpretive exercises and performance assessment tasks require considerable administration time, but a single test item or assessment task is still inadequate for measuring an intended outcome. One exercise

calling for the interpretation of graphs, for example, is not sufficient to measure adequately the ability to interpret graphs. The nature of the data or the type of graph may be the most influential factor in determining whether it is interpreted properly. When several graphs are used, the effect of such factors is minimized, and we obtain a more representative sample of the ability to interpret graphs.

A similar situation occurs with the use of performance assessment tasks. The performance of any single task depends too heavily on the particular sample of information called for by the task, and thus the only feasible solution is to confine each assessment of complex outcomes to a rather limited area (e.g., graph interpretation or problem solving) and to assess more often. In any event, our aim should be to obtain as representative a sample of student performance as possible in each area to be assessed. Other things being equal, the greater the number of test items or assessment tasks, the greater the likelihood of an adequate sample and the more reliable the results.

Eliminating Irrelevant Barriers to the Performance

As was discussed in Chapter 4, the validity of the assessment may be undermined by the influence of construct-irrelevant factors or abilities that are ancillary to the intent of the assessment. Hence, care must be taken when constructing items or assessment tasks to eliminate any construct-irrelevant factors that might prevent students from performing their best. If students have achieved a particular learning outcome (e.g., knowledge of terms), we would want them to answer correctly those test items that measure the attainment of that learning outcome. We would be very unhappy (and so would they) if they answered such test items incorrectly merely because the sentence structure was too complex, the vocabulary too difficult, or the type of response called for unclear. These factors, which are extraneous to the central purpose of the measurement, limit and modify the students' responses and prevent them from showing their true levels of achievement. Such factors are as unfair as determining a person's running ability when an ankle is sprained. Although a measure of running ability would be obtained, the performance would be restricted by a factor we did not intend to include in our measurement.

One way to eliminate factors that are extraneous to the purpose of a measurement is to be certain that all students have the prerequisite skills and abilities needed to make the response. These have been called enabling behaviors because they enable the student to make the response but are not meant to be critical factors in the measurement. That is, they are a necessary but not sufficient condition for responding correctly or performing a task well. Probably the most important enabling behavior in objective testing is reading skill. In written performance assessments, skill in written expression is an additional factor to be considered. In measuring understanding, thinking skills, and other complex learning outcomes, knowledge of certain facts and simple computational skills also might be necessary prerequisites.

In constructing tests and assessments, then, we need to strive for items and tasks that measure achievement of the specific learning outcomes and not differences in ancillary abilities. Differences in reading ability, computational skill, communication skills, and the like should not influence the students' responses unless such outcomes are specifically being measured. The only functional difference between those students who perform well

on a task and those who perform poorly should be the possession of the knowledge, understanding, or other learning outcome being measured by the task. All other differences are extraneous to the purpose of the task, and their influence should be eliminated or controlled for valid results.

A special problem in preventing extraneous factors from distorting our test and assessment results is avoiding ambiguity. Objective test items are especially subject to misinterpretation when long, complex sentences are used; when the vocabulary is unnecessarily difficult; and when words that lack precise meaning are used. Thus, from the viewpoint of both level of reading difficulty and preciseness of meaning, the antidote for ambiguity seems to be a careful choice of words and the use of brief, concise sentences. In some cases, ambiguity can be reduced by using pictures or other illustrative material in place of verbal descriptions. When this is done, the illustrative material must, of course, also be carefully checked to make sure it is clear and unambiguous.

Care should be taken to avoid any racial, ethnic, or gender bias in preparing the test items and performance assessment tasks. The vocabulary and task situations should be acceptable to various racial and ethnic groups and to both males and females and should be free of stereotyping. For example, in presenting characters in a story problem, a reading passage, or other test situation, minorities should not be portrayed in stereotypical fashion. Similarly, task situations should not always place males in such traditional roles as athlete, business executive, and professional person and females in such traditional roles as homemaker, teacher, and nurse. A balanced use of different roles for minorities and males and females is necessary if we are to avoid bias as a possible barrier to maximum performance. See the "Guidelines" box.

Avoiding Unintended Clues in Objective Test Items

Test items should be constructed so that students obtain the correct answer only if they have attained the desired learning outcome. In the preceding principle, we were concerned with those factors that prevent students from responding correctly even though they have attained the desired learning outcome. Here we are concerned with those factors that

GUIDELINES

Possible Barriers in Test Items and Assessment Tasks

- Ambiguous statements
- Excessive wordiness
- Difficult vocabulary
- Complex sentence structure
- Unclear instructions
- Unclear illustrative material
- Racial, ethnic, or gender bias

enable students to respond correctly even though they lack the necessary achievement. These are the clues, some rather obvious and some very subtle, that inadvertently creep into test items during construction. They lead the poor achiever to the correct answer and thereby prevent the items from functioning as intended. When test items are short-circuited in this manner, they provide invalid evidence of achievement. Note how *an* provides a clue to the following item.

EXAMPLE A porpoise is an

A. plant
B. reptile
C. animal
D. bird

Such clues are not limited to selection-type items, as shown in the following supply-type item.

EXAMPLE A piece of land that is completely surrounded by water is known as an _____.

The clue is much less obvious to the person constructing this test item than the clue in our first illustration. To the student taking the test, however, it is readily apparent. The two most plausible answers are *island* and *peninsula*. Because *peninsula* begins with a consonant sound and does not follow the article *an*, it is ruled out as a possibility. This does not imply, of course, that students need to know the rules for good grammatical structure in order to use clues; most clues are analyzed in terms of partial knowledge and hunches. *An peninsula* just does not sound right to the student, so the word *island* is used, and the correct answer is obtained.

Leads to the correct answer may also be provided by simple verbal associations. Note how the word *wind* in the following item provides a clue to the answer.

EXAMPLE Which of the following instruments is used to determine the direction of the wind?

A Anemometer
B Barometer
C Hygrometer
D Wind vane

Rather than lead the uninformed to the correct answer, such clues should lead the poor achiever away from the correct answer. In the following item, the same clue makes

wind vane a plausible (but incorrect) answer for those students who have not learned the uses of the various weather instruments.

EXAMPLE Which one of the following instruments is used to determine the speed of the wind?

A Anemometer
B Barometer
C Hygrometer
(D) Wind vane

Verbal clues need not be as obvious as these. In fact, the clues that appear in the final version of a test are usually rather subtle, as they are based on partial knowledge and verbal associations not readily apparent to the casual observer. For example, at first glance the following item appears to be free from clues.

EXAMPLE Which one of the following is used to prevent polio?

A Gamma globulin
B Penicillin
C Salk vaccine
(D) Sulfa

An examination of this item, however, will indicate that the word *vaccine* provides a clue to the answer. All the student needs to know to answer the item correctly is that a vaccine is used to prevent disease. Because most students have been vaccinated at one time or another, they probably possess this partial knowledge needed to make the clue apparent to them. Some students also may have developed a verbal association between *Salk* and *polio* and respond correctly on that basis. In either case, partial knowledge can lead to the correct answer and prevent the item from functioning as intended.

Another type of subtle clue is one based on the words used to qualify statements. For example, true–false statements that include qualifiers such as *sometimes, usually, generally*, and the like are most often true, whereas statements containing absolutes such as *always, never, none*, and *only* are most often false. Such words have been called specific determiners. They are difficult to remove from true–false items because true statements generally must be qualified, and false statements frequently must be stated in absolute terms to make them clearly false.

Other common clues in selection-type items include (a) stating correct answers in textbook language or in greater detail than incorrect answers, (b) making correct answers longer than incorrect answers, and (c) placing the correct answers in some identifiable pattern (e.g., T, F, T, F). Some of these clues are more likely to be detected by low-achieving students who are desperately searching for some basis for answering. See the "Guidelines" box.

GUIDELINES

Common Clues in Test Items

- Grammatical inconsistencies
- Verbal associations
- Specific determiners (e.g., *always*)
- Phrasing of correct responses
- Length of correct responses
- Location of correct responses

General Suggestions for Writing Test Items and Assessment Tasks

In preparing a set of test items or assessment tasks, the following general rules apply to all types of items or tasks. The specific rules for writing each item or task type are described and illustrated in the following chapters.

1. **Use your test and assessment specifications as a guide.** The specifications describe the performance to be measured and the sample of learning outcomes to measure. Thus, they serve as an aid for selecting the types of items and tasks to prepare, for writing, and for determining how many items and tasks are needed for each subdomain of achievement.

2. **Write more items and tasks than needed.** Preparing more test items and assessment tasks than needed will permit the weaker items and tasks to be discarded during later review. It will also make it easier to match the final set of items and tasks to the specifications.

3. **Write the items and tasks well in advance of the testing date.** Setting the items and tasks aside for several days and then reviewing them with a fresh outlook will reveal any lack of clarity or ambiguity that was overlooked during their preparation. It is frequently surprising how many defects slipped through during the original writing.

4. **Write each test item and assessment task so that the task to be performed is clearly defined and it calls forth the performance described in the intended learning outcome.** Clarity is obtained by carefully formulating the question and instructions, using simple and direct language, using correct punctuation and grammar, and avoiding unnecessary wording. During both writing and review, compare the task students are asked to perform to the learning outcome the task is designed to measure to make sure the two match.

5. **Write each item or task at an appropriate reading level.** Keep the reading difficulty and vocabulary level as simple as possible to prevent these factors from distorting the results. Students' responses should be determined by the performance being measured, not by some factor the item or task was not designed to measure.

6. **Write each item or task so that it does not provide help in responding to other items or tasks.** Unless care is taken during writing, one item may provide information that is

useful in answering another item. For example, a name, date, or fact inadvertently included in the stem of a multiple-choice item may be called for in a short-answer item in another part of the test.

7. Write each item so that the answer is one that would be agreed on by experts or, in the case of assessment tasks, the responses judged excellent would be agreed on by experts. This rule is easy to satisfy when measuring factual knowledge but more difficult when measuring complex outcomes calling for the extended essays or other types of performance.

8. Whenever a test item or assessment task is revised, recheck its relevance. When reviewing items or tasks for appropriateness, clarity, difficulty, and freedom from clues and bias, some revision is often needed. After revising an item or task, check to be sure that it still provides a relevant measure of the intended learning outcome. Even slight changes can sometimes modify the function of an item or a task.

Focusing on Improving Learning and Instruction

The ultimate purpose of testing and assessment is to improve student learning. As you construct classroom tests and assessments, keep in mind the extent to which it is likely to contribute, directly or indirectly, toward this end. Well-constructed classroom tests and assessments should increase both the quantity and the quality of student learning.

1. Tests and assessments can have a desirable influence on student learning if teachers pay attention to the breadth and depth of content and learning outcomes measured. When we select a representative sample of content from all the areas covered in our instruction, we are emphasizing to the students that they must devote attention to all areas. They cannot neglect some aspects of the course and do well on the tests. By giving more weight to high-priority objectives that are emphasized in instruction, we also encourage at the same time a concentration of effort on those objectives thought to be most important. Similarly, when our tests measure a variety of types of learning outcomes, the students soon learn that a mass of memorized factual information is not sufficient. They must also learn to interpret and apply facts, develop deep conceptual understandings, draw conclusions, recognize assumptions, identify cause-and-effect relations, generate hypotheses, solve meaningful problems, and the like. This discourages the students from depending solely on memorization as a basis for learning and encourages them to develop the use of more complex mental processes.

2. Constructing tests and assessments that measure a variety of learning outcomes should also lead to improved teaching procedures and, thus, indirectly to improved student learning. As we translate the various learning outcomes into test items and assessment tasks, we clarify our notion of understandings, thinking skills, and other complex learning outcomes. This clarification enables us to plan the learning experiences of students more effectively and increases the degree to which we emphasize understandings, thinking skills, and other complex learning outcomes in our teaching. A well-constructed test or assessment frequently leads to a review of teaching procedures and to the abandonment of those that encourage rote learning.

3. Tests and assessments will contribute to improved teacher–student relations (with a beneficial effect on student learning) if students view the tests and assessments as fair and useful measures of their achievement. We can make fairness apparent by including a representative sample of the learning outcomes that have been emphasized during instruction by writing clear directions, by making certain that the intent of each item or task is clear and that each item is free of any type of bias that would prevent a knowledgeable person from performing well, and by providing adequate time limits. Student recognition of usefulness, however, depends as much on what we do with the results of the test or assessment as on the characteristics of the instrument itself. We make the usefulness apparent by using the results as a basis for guiding and improving learning.

SUMMARY

Planning the classroom tests and assessments involves (a) determining the purpose of measurement, (b) developing a set of specifications, (c) selecting appropriate types of test items and assessment tasks, and (d) preparing a set of relevant items and tasks.

Classroom tests and assessments can be used for a variety of instructional purposes, best described in terms of their location in the instructional sequence. There are (a) pretests at the beginning of a course or unit to determine learning readiness, to aid in instructional planning, and to make advanced placements; (b) tests and assessments during instruction to improve and direct student learning and to identify and remedy learning errors; and (c) end-of-instruction tests and assessments used at the end of a course or unit to assign grades, certify accomplishment, or evaluate teaching. Each of these types of classroom testing and assessment places different demands on sampling of items and tasks and the type of interpretation used (i.e., criterion referenced or norm referenced).

A sample of student performance is more likely to be representative if a set of specifications is used in planning the test or assessment. Specifications define and delimit the achievement domain to be measured and describe the sample of test items and assessment tasks to be prepared. One form of specifications is a two-way chart called a table of specifications. Building the table involves (a) obtaining the list of instructional objectives, (b) outlining the course content, and (c) preparing the two-way chart that relates the instructional objectives to the course content and specifies the nature of the desired sample of items and tasks. Although a table of specifications is especially useful in preparing summative tests and assessments (because of the broad coverage), it is also useful in preparing some formative tests and assessments. In other cases, however, a test or assessment plan might be limited to a brief list of specific and precisely stated learning outcomes, or it might contain a comprehensive and detailed set of specifications with illustrative sample items or tasks.

The major categories of objective tests or performance assessments may be further subdivided into the following basic types of test items and assessment tasks.

Objective Test

A. Supply type
1. Short answer
2. Completion

B. Selection type
1. True–false or alternative response
2. Matching
3. Multiple choice

Performance Assessment
A. Extended response
B. Restricted response

Objective tests present students with a highly structured task that limits their response to supplying a word, brief phrase, number, or symbol or to selecting the answer from among a given number of alternatives. Performance assessments permit students to respond by selecting, organizing, and presenting ideas or performing in a way they consider appropriate. Both objective tests and performance assessments serve useful purposes in measuring student achievement. The type to use in a particular situation is best determined by the learning outcomes to be measured and by the unique advantages and limitations of each approach. A good practice is to include both objective test items and performance assessment tasks in a comprehensive measurement of student achievement.

The preparation of a set of relevant test items and assessment tasks involves (a) matching the items and tasks to the learning outcomes as directly as possible, (b) obtaining a representative sample of all intended outcomes, (c) eliminating irrelevant barriers to the answer, (d) preventing unintended clues to the response, and (e) focusing on improving learning and instruction. The rules for constructing each type of objective test item and performance assessment task are described in the chapters that follow.

LEARNING EXERCISES

1. Describe the nature of a readiness pretest and how the results might be used in teaching.
2. Describe how formative tests and assessments and summative tests and assessments differ.
3. What are the advantages of using a two-way chart when preparing specifications for tests and assessments? For what type of testing and assessment is it most useful? Why?
4. Why is it important for classroom tests and assessments to measure a representative sample of intended learning outcomes?
5. What types of information should be considered during each of the following steps in constructing tests and assessments?
   a. Developing a set of specifications
   b. Selecting the types of test items and assessment tasks to use
   c. Writing the test items and assessment tasks
6. List several learning outcomes that are best measured with objective test items. List several that require the use of performance assessment tasks.
7. List as many specific factors as you can think of that might prevent some students from performing their best even though they possessed the knowledge and skills the item or task was designed to measure.
8. List as many specific factors as you can think of that would enable some students to answer an objective test item correctly even though they lacked the knowledge the item was designed to measure.
9. Assume that you are going to prepare a brief unit test for a unit of work in a course in your major teaching area. How would you proceed? How would your procedure differ if it were to be an end-of-course summative test or assessment?

REFERENCES

National Council of Teachers of Mathematics. (1989). *Curriculum and evaluation standards for school mathematics.* Reston, VA: National Council of Teachers of Mathematics.

FURTHER READING

Gronlund, N. E. (2005). *Assessment of student achievement* (8th ed.). Boston: Allyn & Bacon. Chapter 3 provides guidelines on planning achievement tests.

Millman, J., & Greene, J. (1989). The specification and development of tests of achievement and ability. In R. L. Linn (Ed.), *Educational measurement* (3rd ed.) New York: Macmillan. A comprehensive and advanced treatment of test development.

Popham, W. J. (2007). *Classroom assessment: What teachers need to know* (5th ed.). Boston: Allyn & Bacon. Chapter 5 discusses issues of what to assess and how to assess it.