

Assessing Student Performance



This Chapter Provides Information That Will Help You To:

1. Describe the role that classroom assessment plays in data-driven decision making.
2. Distinguish the purposes and procedures for preassessment, formative assessment, and summative assessment.
3. Adjust assessments for English language learners and students with special needs.
4. Establish a framework for effective classroom assessment.
5. Select ways to conduct performance-based evaluation.
6. Prepare teacher-made tests using the guidelines provided.
7. Fulfill the teacher's role in standardized testing.

CLASSROOM ASSESSMENT

What Is Classroom Assessment?

Data-Driven Decision Making

Types of Assessments for Decision Making

Norm-Referenced and Criterion-Referenced Evaluation

Characteristics of Good Assessment Instruments

Adapting Assessments for English Language Learners

Adapting Assessments for Students with Special Needs

ESTABLISHING A FRAMEWORK FOR ASSESSMENT

PERFORMANCE-BASED ASSESSMENT

Product Assessments

Performance Assessments

Ways to Rate Student Products or Performances

TEACHER-MADE TESTS

Planning the Classroom Test

Selecting and Preparing Test Questions

Assembling the Test

Administering the Test

Scoring the Test

ACHIEVEMENT TESTS

Types of Achievement Test Scores

The Teacher's Role in Standardized Testing

Assessing students is necessary to help teachers determine the degree to which educational objectives have been achieved and to help teachers know their students as individuals. As you consider student assessment, think about these questions: What is classroom assessment? How do teachers develop an overall plan for assessment in their classrooms? How can performance-based assessments be conducted? How can teacher-made tests be prepared and administered? What are achievement tests, and what is the teacher's role in standardized testing? These questions will be explored in this chapter.

Classroom Assessment

Teachers cannot walk into a classroom cold and plan to address assessment later in the unit or at the marking period. Issues of classroom assessment must be considered from the very beginning moments of planning for instruction. This section addresses classroom assessment, data-driven decision making, types of assessments, characteristics of good assessment instruments, and ways to adapt assessments for English language learners and students with special needs.

WHAT IS CLASSROOM ASSESSMENT?

Classroom assessment can be defined as the collection, evaluation, and use of information to help teachers make decisions that improve student learning (Brookhart & Nitko, 2015; McMillan, 2014; Waugh & Gronlund, 2013). It is more than testing and measurement. There are four components to implementing classroom assessment:

1. *Purpose.* The first step in any assessment is to clarify the specific purposes of gathering information. Why am I doing this assessment? What teacher decision making is enhanced by the information being gathered? There are many possible purposes, such as:

- To diagnosis students' strengths and weaknesses
- To monitor student progress toward achieving objectives
- To assign grades
- To provide students with feedback
- To motivate students

Knowing the reason for the assessment is important because it will determine what the assessment will look like, how it is administered and scored, and how the results will be used.

2. *Measurement.* **Measurement** is the process by which traits, characteristics, or behaviors are differentiated (McMillan, 2014). It is a way of assigning numbers to behavior or performance (e.g., 87 percent of correct answers on a test). The process of measurement may be very formal and quantitative, or it may be less formal and even qualitative. Various techniques can be used to measure a learning target, such as tests, observations, and rubrics.

3. *Evaluation.* **Evaluation** involves an interpretation of the data that has been gathered through measurement and the placement of a value on that performance (McMillan, 2014). Teachers' professional judgment is applied in evaluation. Is 87 percent correct on

a test an average score or a high score? Does that score represent an acceptable level of proficiency? Teachers set performance standards to determine whether a performance is “good” or “bad.” Criteria are then identified as a means to gauge whether students meet the performance standard.

4. *Use.* Once measurement and evaluation have taken place, teachers can employ that assessment information for three classroom uses (McMillan, 2014): diagnosis, grading, and instruction. Diagnosis decisions are made about individual students as well as about group strengths, weaknesses, and needs. This information helps the teacher diagnose specific areas that need attention or where progress is being made. Grading decisions are made based on measurement-driven information. Finally, teachers use assessment information to make instructional decisions. Assessment data is needed for data-driven decision making, as we see in the next section.

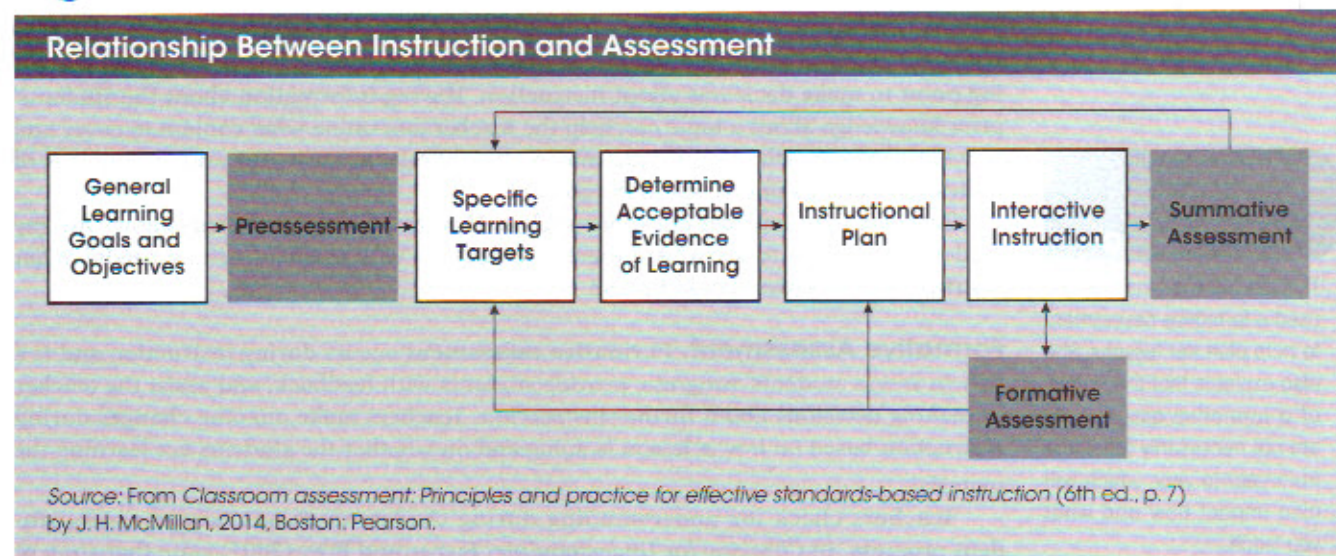
DATA-DRIVEN DECISION MAKING

Teachers make continuous decisions before, during, and after instruction. But they do not make these decisions in a vacuum; they seek out and use much information to guide their decision making. **Data-driven decision making** means that teachers use information from many sources to make decisions about planning, teaching, and assessing.

Preinstructional decisions are needed to set learning goals, select appropriate content and instructional activities, differentiate instruction, and prepare instructional materials. During instruction, teachers make decisions about the delivery and pace of the lesson, keeping students’ attention, controlling students’ behavior, checking for student understanding, and making adjustments in the lesson. After instruction, teachers evaluate student learning, instructional activities, and themselves to know how to grade students, what to teach next, and how to improve.

Assessment is used to provide vital information the teacher needs to make these decisions before, during, and after instruction. The relationship between instruction and assessment is displayed in Figure 11.1. That figure also highlights the three types of assessments discussed in the next section.

Figure 11.1





VOICES from the Classroom

CLAUDIA ARGUELLO COCA, fifth-grade teacher, Las Cruces, New Mexico

DATA-DRIVEN DECISION MAKING

In my classroom, data drives my instruction in every subject area. To serve my students at the highest level possible, I need to know what they know, what they don't know, and what their misconceptions are about core content areas.

I use homogenous student grouping when providing guided reading instruction so that I can offer very specific strategies to those students who are struggling. I create my groups by ongoing assessments that are teacher created, by New Mexico's standards-based assessment (NMSBA) scores, and by our district measures of academic progress (MAPS) scores. All of these scores are used to create a picture of each of my students as a learner.

I use ongoing assessments to make decisions about student progress during unit studies. I use formal assessments and language objectives to measure students' proficiency. By using all of these tools, I am able to create fluid groups for instruction, reteaching, and enrichment. Using these assessments also helps when grouping students for daily response to intervention (RTI) instruction.



TYPES OF ASSESSMENTS FOR DECISION MAKING

There are three types of assessment, each serving a different purpose and each being conducted at different times (Brookhart & Nitko, 2015; McMillan, 2014). Each type of assessment is linked to a stage of instruction—before instruction, during instruction, and after instruction.

Preassessment. Preassessment is done by the teacher before instruction to determine students' knowledge, attitudes, and interests. This information is then used as a starting point to make decisions about instruction. Having information about the students' prior knowledge about a topic can help the teacher determine what content to cover and to what depth. It will also provide information about the strengths and weaknesses of individual students in relation to the content.

Teachers may consider questions such as How much do my students know? Are my students motivated to learn? Are there any students with special learning needs? What instructional activities should I plan?

Formative Assessment. Formative assessment occurs during instruction and is a way to assess students' progress, provide students with feedback, and assist the teacher in making decisions about further instruction. Teachers make ongoing changes during instruction based on how a lesson is going and on whether the students are learning the material.

Teachers check for understanding during the lesson as one way to assess student progress. In *Checking for Understanding*, Fisher and Frey (2007) state that student

Formative Assessment and Planning

This teacher explains how she used a formative assessment to help plan her lessons. She also explains that the results of a formative assessment are not necessarily the same from year to year, which will also impact how and what you teach.

understanding can be determined by many formative assessment approaches, such as through oral language, questions, writing, projects and performances, and tests. Marzano (2010) maintains that several approaches to formative assessment must be used to effectively track student progress and adjust instruction.

Tomlinson (2014a) sees formative assessment as a bridge between today's lesson and tomorrow's. She suggests helping students understand the role of formative assessment; focusing on what students know, understand, and need to do; making room for student differences; providing instructive feedback; making feedback user-friendly; assessing persistently; engaging students with formative assessment; looking for patterns; and planning instruction around content requirements and student needs.

Summative Assessment. Summative assessment occurs after instruction and serves as a means to document what students know, understand, and can do. Summative assessment takes place at the end of a unit, marking period, or course and is intended to (1) determine the extent of the students' achievement and competence, (2) provide a basis for assigning grades, and (3) provide the data from which reports to families and transcripts can be prepared. Summative assessment typically examines the student's broad ability as compared to the detailed examination of each component in formative assessment.

NORM-REFERENCED AND CRITERION-REFERENCED EVALUATION

Norm-referenced evaluation is used to interpret a score of an individual by comparing it with the scores of other individuals. Ranking students is the primary issue. For example, a teacher would use a norm-referenced test to compare the reading abilities of students before forming reading groups. A norm-referenced test finds a student's performance level in relation to the levels of other students on the same test. Many standardized achievement and aptitude tests are norm-referenced instruments.

In contrast, **criterion-referenced evaluation** is used to interpret an individual's performance by comparing it to some specified criterion, such as a performance standard. Individual mastery is the primary issue. The object is to determine whether the student can or cannot perform at a certain standard. How well others perform on the same standard is not considered.



what WOULD YOU DECIDE?

YOUR STUDENTS' PRIOR KNOWLEDGE

Students in your class will likely have a wide range of prior knowledge about the content in the next unit you will teach. Most students may know a lot about the topic, or only a little. A few students may have considerable information due to their own life experiences and investigations. It is helpful to identify what they know.

1. To more effectively plan content and experiences in the unit, how can you determine the students' prior knowledge?
2. How might you enable students to contribute to the planning of the unit?
3. How might the grade level that you teach affect your responses to these questions?



VOICES from the Classroom

DANIEL HRDINA, middle school math teacher, Union, New Jersey

ANALYZING TEST RESULTS TO GUIDE PLANNING AND INSTRUCTION

Students are assessed frequently throughout the school year. Oftentimes, the assessments are graded, recorded, quickly reviewed, and placed into a portfolio never to be seen again. Most schools use quarterly benchmarks to track their students' progress through the school year. This process can be beneficial if each skill is examined independently.

For example, this past school year, every seventh-grade student in my middle school took part in a midyear proficiency assessment. When the test was completed, teachers sat down and looked at each question on their students' exams. They categorized

each question under one of the core curriculum content areas. Teachers then looked at each question that every student answered incorrectly and determined if there was a pattern of student mistakes. This enabled teachers to create lesson plans for the following year that would address these weak areas. They also used this information to address areas of concern with their current students for the remainder of the school year.



CHARACTERISTICS OF GOOD ASSESSMENT INSTRUMENTS

The effectiveness of any measuring device depends on its validity, reliability, and practicality (Russell & Airasian, 2012; Waugh & Gronlund, 2013):

1. Validity deals with the extent to which a measuring device measures what it purports to measure. While there are several types of validity, teachers are usually most concerned with content validity. *Content validity* refers to the degree to which an instrument samples the subject matter in the area to be measured or the degree to which it coincides with the instructional objectives to be measured. This is typically the only type of validity used in teacher-made tests and other measures. To determine content validity, you examine the outline of instructional objectives and check the congruence with the test questions. High content validity means that there is a good match between test questions and instructional objectives covered in class. Low content validity means that there is a poor match.

2. Reliability deals with the consistency of results. The more consistent the results, the higher the reliability of the test. If you administer a reliable test and shortly after readminister an equivalent form of the same test, students should score roughly the same as they did on the first test.

All measuring devices have some type of error. **Measurement error** may come from errors in the instrument, errors in the use of the instrument, and errors emanating from the subject's response. The more objective the scoring of a test, the more reliable the score. Factors that may act to make a test less reliable are the length of a test (i.e., very short tests are not very reliable), the ambiguity of directions and questions, and an overload of questions that are too difficult or too easy. Variability in the students' responses can be affected by temporal factors such as poor motivation, lack of interest, improper test environment, poor emotional set, and illness.

3. Practicality refers to the ease of administering the measuring device, the time required, the energy expended to collect the data, and the ease with which the data can be interpreted. For instance, an essay test may be easily prepared by the teacher, but the time required to grade it can make such a test impractical. Conversely, an objective test can be graded relatively quickly but may take much longer to prepare.

In addition to reliability, validity, and practicality, Reeves (2011) maintains that good assessment instruments must address accuracy, fairness, specificity, and timeliness. Without those qualities, assessments fall short of effectively serving their purposes. Furthermore, Reeves suggests that the selection of assessments take into account the amount of time teachers would need to complete the assessment and grading, and he offers some time-saving strategies for teachers concerning assessments.

ADAPTING ASSESSMENTS FOR ENGLISH LANGUAGE LEARNERS

English language learners (ELLs) may know the content on which they are being assessed but have difficulty demonstrating their knowledge with standard assessments. These challenges often stem from the vocabulary, reading, and writing that are part of a typical assessment. ELLs may experience difficulties in assessments due to (1) difficulty in comprehension of the test language, (2) difficulty in expressing what they know, (3) lack of content and cultural knowledge of the test items, (4) unfamiliarity of test types, and (5) emotional stress (McMillan, 2014). Given this, teachers must adapt assessments to accurately determine the content knowledge of ELLs.

Consider the following suggestions when selecting or adapting assessments for ELLs (Balderrama & Diaz-Rico, 2006; Echevarria, Vogt, & Short, 2013; Rothenberg & Fisher, 2007):

- 1. Variety.** Use a variety of assessment tools and do not rely on only one or two types of assessments to measure student learning.
- 2. Formative assessment.** Formative assessment provides information about students' abilities and performance, and this information can be used to adjust instruction to meet the needs of ELLs.
- 3. Range.** Adapt the number of items that ELLs are expected to complete, such as only the odd or only the even questions.
- 4. Time.** Adapt the amount of time that ELLs have for completing a task, such as providing more processing time or breaking tasks down into manageable chunks. Unless taking a timed test is required, allowing additional time should not affect a student's score.
- 5. Level of support.** Adapt the amount of scaffolding provided to ELLs during assessments by asking an aide, peer assistant, or parent volunteer to read or explain the task or even to read the test items aloud. If you are assessing for content knowledge, then it is acceptable to have someone help with reading or clarifying what is expected in the task.
- 6. Difficulty.** Adapt the skill level, the type of problem or task, and the process for how ELLs can approach the task, such as allowing use of a calculator, dictionary, or simplified directions. None of these adaptations reduces the expectation of what the student should know; it only makes it easier for him or her to demonstrate understanding.
- 7. Product.** Adapt the type of response ELLs are allowed to provide, such as permitting a drawing, a hands-on demonstration, a verbal response, or a translated response, if necessary. These alternatives do not rely on sophisticated English usage.
- 8. Participation.** Adapt ELLs' degree of active involvement in assessment, such as encouraging individual or cooperative group self-assessment and assistance in creating rubrics. Content learning is enhanced for all students, but especially for ELLs, through interaction and group work. English language learners, particularly those in the upper grades, can be involved in assessing their own progress.



sample

STANDARDS

ASSESSMENT

There are 10 InTASC standards (see pages xviii–xix), and each standard in the original document includes a list of performances, essential knowledge, and critical dispositions to indicate more clearly what is intended in the standard.

Because this chapter deals with assessment, some representative statements from InTASC Standard #6, Assessment, are listed here concerning topics in this chapter.

PERFORMANCES

- The teacher balances the use of formative and summative assessment as appropriate to support, verify, and document learning.
- The teacher designs assessments that match learning objectives with assessment methods and minimize sources of bias that can distort assessment results.

ESSENTIAL KNOWLEDGE

- The teacher understands the range of types and multiples purposes of assessment and how to design, adapt, or select appropriate assessments to address specific learning goals and individual differences, and to minimize sources of bias.
- The teacher knows when and how to engage learners in analyzing their own assessment results and in helping to set goals for their own learning.

CRITICAL DISPOSITIONS

- The teacher takes responsibility for aligning instruction and assessment with learning goals.
- The teacher is committed to making accommodations in assessments and testing conditions, especially for learners with disabilities and language learning needs.

9. Feedback. Provide ongoing feedback to students as a result of formal and informal assessments so that they see can see their progress and make adjustments if needed.

Because differentiating instruction is useful in meeting the needs of diverse learners, Tomlinson and Moon (2013) suggest planning for assessment in a differentiated classroom. They state that there are five components of instruction that can be differentiated: content, process, product, affect, and the learning environment. Adjustments can be made in these components based on the learner's readiness, interest, and learning profile. Formative and summative assessment can be conducted based on these components.

ADAPTING ASSESSMENTS FOR STUDENTS WITH SPECIAL NEEDS

Students with special needs or disabilities may have difficulty with standard assessment and test-taking procedures. As a result, teachers often need to make accommodations or changes in formative and summative assessments for these students.

Depending on the type of special need or disability, students may encounter problems in testing situations (McMillan, 2014). These problems include comprehension difficulties, auditory difficulties, visual difficulties, time constraint difficulties, anxiety, embarrassment, and variability in behavior from one day to the next. When considering summative tests, accommodations for these problems can be grouped into the following categories (J. W. Wood, 2006):

1. Test construction. Attention needs to be given to test directions and the test format. When modifying test directions, take the following suggestions into account (Mastropieri & Scruggs, 2014); some of these guidelines can be used for all students. Read written directions aloud, slowly, and give students opportunities to ask questions

about directions. Keep directions short and simple. Give examples of how to answer questions. Focus attention by underlining verbs. Provide separate directions for each section of the test. Check students' understanding of the directions. During the test, observe students' answers to be sure that the students understand the directions. Adjust the reading level of the items or provide assistance with reading if necessary.

The general test format should be designed to simplify the amount of information that is processed at one time. Leave plenty of white space so that students do not feel overwhelmed. Use large printing with adequate space between items. Create clearly distinguished short sections. For multiple choice questions, list the alternatives vertically. Number each page of the test. If possible, design the test format of an adapted test to look as much like the test for other students as possible (Salend, 2009). Other accommodations to the test format may be needed depending on the needs of the learner.

2. *Test administration.* Adaptations during test administration involve changes in procedures that lessen the negative effect of disabilities while the student is taking the test. Most of these procedural accommodations depend on the nature of the disability or difficulty mentioned earlier (e.g., comprehension difficulties, time constraint difficulties). A good rule of thumb is to provide students with disabilities 50 percent additional time to complete a test (Reynolds, Livingston, & Willson, 2009).

3. *Testing site.* It may be necessary to allow students with special needs to take the test in a different location than the regular classroom. The alternative site is often a resource room in the school or another room that is quiet with few distractions. As long as someone can monitor the test, the student will have more opportunities to ask questions and feel less embarrassed when asking questions for clarification or further explanation.

Establishing a Framework for Assessment

Even before instruction begins, you need to establish a framework for student assessment by addressing the following issues:

1. *Identify the reasons for assessing student performance.* A classroom test, for instance, can serve a variety of purposes, such as judging student mastery, measuring growth over time, ranking pupils in terms of their achievement, and diagnosing student difficulties. Different purposes lend themselves to various assessment measures. Therefore, it is essential that you know the major use of the assessment results.

2. *Plan to gather information for both formative and summative assessment.* Formative assessment should be conducted frequently to provide feedback about the student's developing knowledge and skills. In fact, all scores for formative assessment may not need to be recorded for the report card grade. Summative assessment is more comprehensive and is conducted at selected concluding points, such as at the end of a chapter or unit. Scores from summative assessments are tallied as part of the report card grade.

3. *Identify course content that is to be assessed.* This content is often included in a course curriculum guide. It will be translated into instructional objectives for daily lesson plans. These objectives will likely be written for the cognitive, psychomotor, and affective learning domains.

4. *Relate the assessment to the prestated objectives.* Each objective should be evaluated, but the means for gathering assessment information will vary depending on the stated objectives and the purpose of the assessment.

Mapping Curriculum to Standards

Backward design allows you to design a lesson based on the goal you would like to achieve. Your assessments can be designed in parts to help ensure alignment to your goal. Are there certain standards that you must create your lessons around?

5. *Include assessment measures in all three learning domains.* Instructional objectives should be written for each of the three learning domains, and each objective should be assessed. Different types of assessment instruments are often used to measure student performance in each domain. The cognitive domain is often assessed through the use of (a) tests developed by teachers themselves; (b) achievement tests developed by local, regional, or state agencies, which may or may not be normed; and (c) nationally standardized tests and scales. Skills in the psychomotor domain are evaluated through performance and product tests.

Objectives in the affective learning domain are more difficult to assess than objectives in the other domains. Techniques often used include attitude scales, opinion polls, questionnaires, open-ended questions, checklists, observation, group discussion, anecdotal records, interviews, and role-playing. Observation of students in the process of performing a particular task can also provide useful information concerning their attitudes toward the task.

6. *Use multiple approaches when gathering assessment information.* Do not rely only on teacher-made tests. Also use performance-based assessments when appropriate.

7. *Devise an assessment plan before instruction.* This will guide your planning and preparation for the class sessions. Then communicate the objectives and assessment criteria to the students.

Performance-Based Assessment

As an alternative to teacher-made tests, teachers can measure student learning by assessing products that students prepare reflecting their learning or by assessing actual student performances designed to demonstrate their learning. This is **performance-based assessment**. Once the student produces the product or conducts the performance, the teacher then needs to use various types of rating scales or checklists to translate the student performance into some type of score.

Because performance-based assessments are seen as alternatives to teacher-made tests, terms such as *alternative assessment*, *authentic assessment*, and *direct assessment* have been used to describe performance-based assessments. There are two types of performance-based assessments: product assessment and performance assessment (see Table 11.1).

Table 11.1 Performance-Based Assessments

PRODUCTS	PERFORMANCES
Portfolio	Oral presentation
Work sample	Presentation with media
Project or report	Science lab demonstration
Research paper	Athletic demonstration
Science lab report	Musical, dance, or dramatic presentation
Log or journal	Debate
Model	Demonstration of specific tasks
Media products (computer presentation file, audio- or videotape)	Participation in an event
	Interview with the student

PRODUCT ASSESSMENTS

Instead of completing a traditional paper-and-pencil test, students are asked to prepare a product reflecting their learning, and the teacher then completes a rating scale or composes written comments to assess student learning as reflected in the product. Product assessments include (1) directions that outline the nature of the products that students will develop; (2) the product itself, which students prepare; and (3) a rating scale of some kind that assesses the nature of the student product. There are many types of products that students might be asked to prepare. Some of the more common types of student products are discussed here.

Portfolios. A **portfolio** is a collection of a student's work over time that demonstrates mastery of specific performance criteria against which the tasks in the portfolio can be judged. The student systematically chooses and compiles work products that go into the portfolio. Work products can be items such as reports, letters, drawings, journals, and other documents. Portfolios are intended to actively involve students in learning, develop self-assessment skills, and provide insights to students and teachers on progress and accomplishments (Burke, K., 2009; Popham, 2014).

You can guide students through portfolio development by identifying the characteristics of good academic work, applying criteria for good work to their own work, using peer-revision groups to refine and assess their work, and selecting work so that the portfolio creates a portrait of the student as a learner. Teachers who have implemented portfolios report that (1) students have a richer, more positive, and expanded sense of their progress; (2) assessment becomes collaborative rather than competitive; (3) teachers obtain a richer, clearer view of their students over time; and (4) records of what students are actually doing are available to teachers.

Some teachers think portfolios are appropriate for use only in the elementary grades, but teachers at the middle and secondary levels have also used portfolios successfully. Geometry portfolios, for example, might include work that displays problem solving, communication, or other skills. Language arts and social studies portfolios might include writings and other projects that offer evidence of progress and goal setting.

Work Samples. Gathering work samples provides a longitudinal record of student progress. Work samples might include (1) written work such as a report, test, or story; (2) artwork; (3) tape recordings; (4) a construction project done in art, industrial arts, or an appropriate subject; or (5) other types of finished products, depending on the subject area. Using work samples probably occurs more often in the elementary grades, but teachers in certain secondary subject areas may use work samples as well.

Experience Summaries. In so-called experience summaries, students record what they have learned from a particular experience. Teachers and students may do this cooperatively. For instance, after going on a field trip to an art gallery, the class could identify facts and concepts that they learned. This listing would be useful for you to determine the extent of student learning.

Logs or Journals. While experience summaries are used for one event, logs or journals are kept on a continuing basis. Students might keep a journal to record their insights, conclusions, and feelings about their classroom experiences. The class can use a notebook or a chart to log daily progress concerning learning activities. This provides a running account of work in the unit and can be used to review and check on previous plans and decisions as the unit progresses.

Projects, Reports, or Papers. Students might prepare various types of projects, reports, or papers that reflect their learning. These products often involve a number of skills and are developed over more than one class period.



VOICES from the Classroom

KAYE LYNN MAZUREK, high school German teacher, Commerce, Michigan

STUDENTS REFLECT ON THEIR ASSESSMENTS

I believe that all assessments can be formative assessments in my high school German classes. We have one major summative assessment each marking period with five sections—listening, reading, writing, speaking, and culture. I use the summative assessment from one unit as a stepping-stone for the next marking period.

When I return the summative assessment from the unit, I have a score sheet that accompanies it. Following the score sheet is a section that asks students to rewrite the sentences with which they had the most mistakes,

to summarize one of their consistent errors, or to reflect on strongest and weakest sections of their test. The task is very brief. Sometimes students don't even realize they're making a consistent error until they actually have to identify it. Thus, this process helps students reflect on their assessments by identifying their strengths, weakness, and areas where they need to give attention for improvement.



Models. Students can prepare models concerning people, places, or topics being addressed in class. The models may contain information and represent certain aspects of their learning.

Media Products. Working individually or in groups, students are asked in many classrooms to prepare an audiotape, videotape, or computer-assisted presentation concerning some aspect of the curriculum. The final media product will represent certain aspects of student learning.

PERFORMANCE ASSESSMENTS

As with product assessments, teachers complete a rating scale or compose written comments to assess student learning reflected in a performance that students are asked to conduct. Performance assessments include (1) directions that outline the nature of the products that students will develop; (2) the product itself, which students prepare; and (3) a rating scale of some kind that assesses the nature of the student product. There are many types of performances that students can be asked to conduct. Some of the more common ones are discussed here.

Oral or Mediated Presentations. Students can be asked to prepare a report on a selected topic and then present it to the rest of the class. Sometimes these reports include instructional media (e.g., computer-assisted presentations). The teacher may choose to rate the actual presentation as a way to gauge student learning.

Actual Demonstrations. Students can be asked to demonstrate their skill in some type of athletic, musical, dance, or dramatic performance. Skills might also be demonstrated in a science lab, a computer class, or any other class where students' learning is best reflected in actual performance.



Participation in an Event. Sometimes a group of students or the entire class is involved in an activity in which students exhibit certain knowledge and skills through their participation. A student's knowledge and skill in forensics or debate, for instance, can be seen through participation.

Interviews. Interviews and individual conferences can be used to evaluate cognitive skills as well as attitudes and values. Students sometimes express ideas and feelings during interviews that they might not otherwise state.

WAYS TO RATE STUDENT PRODUCTS OR PERFORMANCES

It is easy to say that it is a great idea to gather student products or to ask students to perform in certain ways as a means to demonstrate their learning. However, teachers may find it challenging to specify the actual observable criteria that demonstrate student learning and to determine what the actual rating approaches will be. Some guidelines are provided in this section.

Selecting Rating Criteria. The following guidelines can be used when selecting observable criteria for products or performances (Russell & Airasian, 2012):

1. *Select the performance or product to be assessed and then either perform it yourself or imagine yourself performing it.* This step will help you understand the actual student actions that will need to be completed.
2. *List the important aspects of the performance or product.* The most important behaviors or attributes should be reflected in the rating criteria.
3. *Try to limit the number of performance criteria that can reasonably be observed and assessed.* Each item needs to be assessed, so limit the number of performance criteria to 10–15.
4. *If possible, have a group of teachers think through the criteria as a group.* This step will save time and will help to produce a more complete set of criteria than that produced by any single teacher.
5. *Express the criteria in terms of observable student behaviors or product characteristics.* Be specific when stating the criteria.
6. *Avoid vague and ambiguous words.* Avoid wording such as *correctly*, *appropriately*, or *good* that can result in multiple interpretations.
7. *Arrange the performance criteria in the order in which they are likely to be observed.* This will save time when observing and maintain focus on the performance.
8. *Check for existing performance criteria before constructing your own.* Preexisting criteria might be available for your use.

Rating Approaches. Various approaches can be used to rate student products or performances. The most commonly used approaches are reviewed here:

1. **Rating scales.** A **rating scale** allows a teacher to judge student performance along a continuum, rather than as a dichotomy (e.g., present/not present). The rating scale may be numerical (e.g., 1–4), descriptive (using a range of words to reflect the criteria), or a combination of the two. Instead of merely indicating the presence or absence of a trait, as in a checklist, a rating scale enables you to indicate the status or quality of what is being rated.

A **rubric** is a type of rating scale used to judge student performance on a continuum. Rubrics can be created in many ways to describe the levels of student performance based on the performance criteria. Rubrics often have more thorough ways of describing each performance level than simple rating scales. The thorough descriptions of performance

Writing Rubrics

This teacher explains why she has students participate in the creation of rubrics. Students will have a sense of ownership and responsibility over their work because they know exactly what is expected of them. She also explains the purpose of providing examples of work so that students can see how they will be assessed.

criteria in rubrics provide considerable guidance for the students as they prepare their products or performances. See Table 11.2 and Table 11.3 for examples. Useful guides are also available concerning rubrics, such as *How to Create and Use Rubrics for Formative Assessment and Grading* (Brookhart, 2013).

2. *Checklists.* A **checklist** is a written list of performance criteria. A teacher reviews the student product or performance to determine which criteria are met. A check mark indicates that the criterion was met; if it is not, the item is left blank. A checklist enables you to note quickly and effectively whether a trait or characteristic is present, but it does not permit you to rate the quality or frequency of a particular behavior. When that additional information is needed, you may want to use a rating scale or an anecdotal record.

3. *Anecdotal records.* An **anecdotal record** is a written account of events and behaviors that a teacher has observed concerning a student's product or performance. Only those observations that have special significance and cannot be obtained from other assessment methods should be included in an anecdotal record. Anecdotal records should contain factual descriptions of what happened, when it happened, and under what circumstances it happened. The interpretation of the behavior and the recommended action should be noted separately. Each anecdotal record should contain a record of a single incident.

4. *Participation charts.* A **participation chart** is intended only to indicate the degree of involvement and the type of involvement in classroom discussion or activities. A typical form has students' names on one axis and the degree and frequency of involvement on the other axis. The charts are not intended to indicate why students participate.

Table 11.2 A Sample Rubric Assessing a Research Report

	EXCELLENT	SATISFACTORY	NEEDS IMPROVEMENT
COMPREHENSION OF THE SUBJECT	Writing indicates an excellent understanding of the topic. Reflects use of a range of resources.	Writing indicates satisfactory understanding of the topic. Reflects use of one or more resources.	Writing indicates a limited understanding of the topic. Contains factual errors and may miss critical information.
IDEA DEVELOPMENT	Develops relevant ideas clearly and fully. Details and examples explain and clarify the information.	Develops ideas satisfactorily with details and examples.	Develops ideas incompletely with few or no supporting details.
ORGANIZATION	Organizes information logically in the paragraphs. Includes an effective introduction and ending.	Organizes information in an acceptable order that is easy to understand.	Presents ideas with little or no organization.
LANGUAGE USAGE	Uses lively and descriptive language. Details and examples explain and clarify the information. Varies sentence structure.	Uses some descriptive language. Demonstrates some sense of sentence variety.	Uses limited vocabulary. Uses descriptive, simple sentences.
MECHANICS	Writing shows few errors in basic language conventions.	Limited errors in basic language conventions do not interfere with meaning.	Shows many errors in conventions but still conveys some meaning.

Table 11.3 Sample Rubric Rating Scales

4	3	2	1
Excellent	Good	Fair	Poor
Excellent	Above average	Average	Needs work
Advanced	Proficient	Apprentice	Novice
Superior	Advanced	Intermediate	Novice
Consistently	Usually	Sometimes	Rarely
Fully developed	Adequate	Developing	Absent
Exemplary	Mastery	Approaching	Emerging

©Pearson Education

Teacher-Made Tests

To achieve the purpose of teacher-made tests, you need to carefully plan the test and then select and prepare test questions. Next, you need to assemble, administer, and score the test.

PLANNING THE CLASSROOM TEST

The preparation of a classroom test involves many decisions. Advance attention to the following decision areas will likely result in an improved test (Brookhart & Nitko, 2015; McMillan, 2014; Waugh & Gronlund, 2103):

1. *Decide the purpose of the test.* To be helpful, classroom tests must relate to the teacher's instructional objectives and procedures. Classroom achievement tests serve a variety of purposes, such as (a) judging students' mastery of certain knowledge and skills, (b) measuring growth over time, (c) ranking students in terms of their achievement, (d) diagnosing student difficulties, (e) determining the effectiveness of the curriculum, (f) encouraging good study habits, and (g) motivating students. These purposes are not mutually exclusive.

2. *Decide what is to be tested.* What should be tested? The answer depends on the instructional objectives and what has been stressed in class. As noted later in this discussion, the table of specifications should be determined prior to instruction, and this should serve as the basis for preparing a test. The table of specifications includes instructional objectives and content. The relative emphasis on the three domains of learning should be reflected in the table of specifications. Furthermore, an effective test should include only questions that address important outcomes of instruction.

3. *Prepare a table of specifications showing the content and objectives to be tested.* A **table of specifications**, sometimes referred to as a test blueprint, identifies the intended outcomes of instruction; it specifies objectives or behaviors that students should attain by the end of the instruction and the content related to these objectives. These statements should be precisely stated to clearly convey intentions. Summative evaluation, then, is intended to examine the degree of success with which the objectives were achieved by the students.

When making a test, the table of specifications is intended to identify the scope and emphasis of the test, to relate the objectives to the content, and to construct a balanced test. The table of specifications should be prepared before instruction. Some teachers

prepare a one-way table of specifications that includes a list of the objectives to be addressed on the test and a notation for the percentage of test questions for each objective. Objectives that are more important or more involved may require a higher percentage of the test questions.

A two-way table of specifications is more descriptive. First, teachers need to prepare a list of the instructional objectives with the level of the learning domain being identified for each. Next, teachers need to outline the course content. Then a two-way chart is prepared that relates the instructional objectives to the instructional content. The table of specifications displayed in Table 11.4 indicates the number of test questions in each learning domain for each content area in a sample unit in science.

4. *Decide when to test.* The use of the test results will determine the frequency of testing. If used in formative evaluation, students should be given short, criterion-referenced tests frequently to keep you apprised of their performance and identify those in need of additional or remedial instruction. Summative evaluation will be conducted less often.

To some extent, the frequency of testing is affected by institutional regulations on marking and reporting grades to parents. More frequent evaluation is superior to less frequent evaluation. Teachers in secondary schools sometimes communicate with each other in an effort to stagger the days when students are tested so that several tests are not scheduled for the same day.

5. *Decide what kinds of questions to use.* There basically are two item formats: essay and objective (multiple choice, matching, true-false). Some item formats are less appropriate than others for measuring certain objectives. Therefore, carefully consider the advantages and disadvantages of each item format. These will be discussed in greater detail in a later section.

Table 11.4 Table of Specifications for a U.S. Constitution Unit in a Middle School Class

OBJECTIVES CONTENT OUTLINE	OBJECTIVES IN THE LEVELS OF THE COGNITIVE DOMAIN						TOTAL NUMBER OF ITEMS	PERCENT OF ITEMS
	1. REMEMBERING	2. UNDERSTANDING	3. APPLYING	4. ANALYZING	5. EVALUATING	6. CREATING		
Executive branch	1	2	4	4	1		12	24
Legislative branch		1	1	4	2	1	9	18
Judicial branch		2	2	2	2	1	9	18
States rights	1	1	2	1	1	1	7	14
Amendments	1	2	5	3	1	1	13	26
Total number of items	3	8	14	14	7	4	50	
Percentage of items	6	16	28	28	14	8		100



what WOULD YOU DECIDE?

USING A TABLE OF SPECIFICATIONS

When planning a unit, your preparation of a table of specifications can help you identify instructional objectives in all appropriate learning domains for each content area within the unit.

1. When you prepare a test, how might you use the table of specifications to ensure that there are test questions covering all the learning domains in all the content areas in the unit?
2. How might this table help you distribute points for the various sections of the test?

6. *Decide how many questions to include in the test.* The length of a test will vary according to its purpose, the kinds of items used, the reliability desired, the length of testing time available, and the age and ability of the students. Formative tests are more in-depth and frequent than summative tests. As explained later, the time needed to complete different item formats varies. The longer the test, the more reliable it tends to be, but only so many items can be completed within a 50-minute class period. The age and ability of the students also must be considered.

7. *Decide the level of difficulty for questions included in the test.* The difficulty of a test should depend to a large extent on its purpose. Item difficulty will not be a factor in the selection of test items if the test results are to be used to describe the status of individual students. However, test difficulty is a consideration when the test results will be used to differentiate among students in terms of their achievement.

8. *Decide the format of the test.* You need to decide about the mechanical and formatting features of the test, such as grouping and arranging test items, directions for answering questions, distribution of correct responses, the layout of items, and other aspects. These will be discussed in a later section.

SELECTING AND PREPARING TEST QUESTIONS

There are two categories of test questions: essay and objective. **Essay questions** include short-answer or restricted responses (about a half page), discussion or extended responses (about two to three pages), or oral responses. **Objective questions** include multiple choice, matching, true-false, and short answer (a single word or several words). Because some item formats are less appropriate than others for measuring certain objectives, teachers should carefully consider the advantages of each type of item format.

Test questions are sometimes included in the teacher's edition of a textbook, and you might consider using them. Carefully examine test items included in the teacher's edition to be sure they are worded effectively, are at the proper levels of the learning domains being tested, are directed at content that was covered during instruction, and are at the appropriate reading level for your students.

Essay Questions. Two common types of essay questions are the restricted response and the extended response. **Restricted response essay questions** establish limits for the form and scope of the students' answers. Students are often asked to discuss one specific aspect of an issue and to limit their answers to perhaps half a page. Restricted essay questions are useful for measuring comprehension, application, and analysis (Stiggins & Chappuis, 2012).

Extended response essay questions, by contrast, place no limits on the student concerning the points to discuss or the organization used in the answer. Extended response questions require students to call on factual knowledge, evaluate this knowledge, organize the ideas, and present these ideas in a logical, coherent way. Skills in synthesis and evaluation are effectively tested with this type of question.

A variety of mental processes can be evaluated through the use of essay tests. From simple to higher-order skills, these include (1) what, who, when, which, and where; (2) list; (3) outline; (4) describe; (5) contrast; (6) compare; (7) explain; (8) discuss; (9) develop; (10) summarize; and (11) evaluate.

There are advantages and disadvantages to the use of essay questions. Advantages include that they are comparatively easier and less time consuming to prepare than objective tests, require that students supply the answer rather than simply select the answer, measure higher-order thinking, and stimulate creativity and freedom of expression.

But essay questions also have several disadvantages. First, they generally sample content in a limited way. Three essay questions may not sample unit content as thoroughly as 50 multiple-choice questions. Therefore, essay tests that have several questions requiring short answers are preferable to a test that has few questions requiring lengthy answers. Second, essay questions are vulnerable to unreliable scoring. This problem of reader reliability can be minimized by careful construction of the test question and by setting up specified scoring procedures. Third, a student may not always understand the question and therefore may be unsure of how to respond, or he or she may be unable to put all appropriate ideas on paper in the given time. Fourth, essay questions require a great deal of time to read and score.

Writing Essay Questions. To minimize the associated problems, it is important to use the following guidelines to carefully prepare essay questions:

1. *Reserve adequate time to prepare test questions.* Allow sufficient time to thoughtfully prepare each question and to be certain that each measures the intended objective, is worded in a simple and clear manner, and is reasonable and can be answered by the students.

2. *Precisely define the direction and scope of the desired response.* This can be done by (a) delimiting the area covered by the question; (b) using descriptive, precise words in the question, such as *outline*, *illustrate*, *define*, and *summarize* (as compared to the term *discuss*); (c) guiding the student into a certain direction by indicating what should be considered and presented in the answer; and (d) indicating the length of the answer (typically one-half page to two pages per question).

3. *Indicate the point value of each question.* Students need to know the point value of each question to be able to decide how much time to spend on it.

4. *Use essay questions to measure objectives that cannot be measured as well with other item formats.* As compared to other item formats, essay questions are especially useful in assessing students' understanding in the upper levels of the cognitive domain, such as application, analysis, synthesis, and evaluation.

5. *Use many questions requiring relatively brief answers, as compared to just a few questions involving long answers.* Brief answers (about one-half page) provide for a broader sampling of content, tend to discourage bias when some teachers grade for quantity rather than quality, enable the teacher to read the answers more rapidly and reliably, and also enable the teacher to direct the student to the desired response area.

6. *Take into consideration the time needed for an adequate response and the time available for the testing period.* Students need to have an appropriate amount of time to write out their responses to all questions.

7. *Adapt the length of response and the complexity of the answer to the maturity of the students.* The depth and breadth of the responses from fifth graders certainly would be different from responses expected from tenth graders.

8. *Do not provide optional questions on an essay test.* Providing options for students undermines the basic function of summative tests (e.g., comparing the performance of students in a common area). Furthermore, it is difficult to construct questions of equal difficulty.

9. *Decide in advance how the essay questions will be scored.* In addition to the substance of student answers, decide how grammar, spelling, punctuation, handwriting, composition, and clarity of expression will be scored.

10. *Prepare a scoring key.* When preparing test questions, also identify what topics should be included in student responses to each essay question and then determine relative points for weak, average, and strong answers.

Evaluating Essay Questions. The usefulness of a carefully constructed essay test can be undermined by improper grading procedures and standards. Essay tests must be reliably graded with either the analytical method or the rating method. In the *analytical method*, the teacher identifies all the specific points or topics that should be included in an appropriate answer. The student's score is determined by the number of points or topics he or she includes in the response to the essay question.

The *rating method* also involves the teacher identifying the points to be included in the answer, but the wholeness of the response is emphasized. Several student papers are selected to reflect the range of responses on a given question. These papers serve as anchor points as the other papers are read and rated. Papers may be read using a 2-point scale (acceptable-unacceptable) up to a 5-point scale (with steps going from superior to inferior). With the 5-point scale, the teacher will place students' papers in one of these five piles upon reading the students' responses. It is preferable that each paper be read and rated twice. The rating method is effective when a large number of essays are to be read.

When evaluating essay questions, consider the following guidelines:

1. *Check the scoring key against actual responses.* Before scoring actually begins, read a few randomly selected papers to determine the appropriateness of the scoring key. Adjustments might be made in the key based on actual responses. The key should not be changed once actual scoring begins.

2. *If spelling, penmanship, grammar, and writing style of the responses are to be scored, it should be done independently of the subject matter content in the responses.*

3. *Score the students' responses anonymously.* If possible, in an effort to reduce biases, conceal the student's name as each paper is evaluated.

4. *Evaluate one question at a time for all students.* This is designed to eliminate the so-called halo effect, a general impression of high, medium, or low quality that might carry over from one response to another for each student. This practice also enables the teacher to more easily keep in mind the standards that are being applied for a particular question.

5. *Shuffle test papers before grading each question.* The ratings given to preceding papers sometimes affect the way teachers rate the next paper to be evaluated. Therefore, shuffle test papers before grading the next test question to try to minimize the effects of the preceding grades.

6. *Try to score all responses to a particular question without interruption.* To keep the standards clearly in mind, the papers should be read without excessive interruption or delay.

7. *Whenever possible, conduct two readings of the test and use the average as the final score.* Two independent readings simply help improve the reliability of the test score. Of course, this may not be realistic at the secondary level if you have 125 student papers to evaluate.



what WOULD YOU DECIDE?

ESSAY QUESTIONS

For middle school and high school teachers who have many students, including essay questions on a test can be very time consuming to prepare the questions and to assess the student responses. Similarly, elementary teachers have many subjects to address and including essay questions in most subject areas may not be practical. Yet essay questions can assess certain knowledge and skills.

1. What are the benefits of essay questions over other types of questions?
2. What could you do to be more time efficient in your evaluation of essay questions that you prepare?

8. *Provide comments and indications of correct answers.* By providing comments, you give students more information about their progress, motivate them, indirectly teach them, and also help to explain the system of grading and determining scores.

9. *Set realistic standards.* Avoid being overly generous or overly demanding as you review student responses.

Objective Questions. The most common types of objective test items are multiple choice, matching, true-false, and short answer.

Multiple Choice. A multiple-choice question has two parts: (1) a *stem*, which contains the problem, and (2) a list of suggested answers, or *responses*. The stem is typically in the form of a question or an incomplete statement. The incorrect responses are often called *distracters*. Generally, four or five responses are listed, and all but one is a distracter.

From the list of responses provided, the student selects the one that is correct or best. Some questions have only one possible answer that is correct. In other questions, students are asked to identify the most appropriate, or best, answer from the choices given.

The *direct question format* has several advantages. It forces the teacher to state the problem clearly in the stem, it reduces the possibility of giving the student grammatical clues, and it may be more easily handled by younger and less able students because less demand is placed on good reading skills. The *incomplete statement format* can also be effective. In this case, the stem must be clear and meaningful and not lead into a series of unrelated true-false statements. A reverse (or negative) version of the direct question format and the incomplete statement format asks the student to select the one incorrect choice from a list of correct choices.

There are several advantages of multiple-choice questions. They (1) can test students in several levels of the cognitive domain, (2) can be scored quickly and accurately, (3) are relatively efficient in terms of the number of questions that can be asked in a given amount of time and the space needed to present the answers, (4) can test a wide range of topics in a short time, and (5) are not significantly affected by guessing.

Multiple-choice questions also have several disadvantages. First, they are very difficult to construct, and plausible-sounding distracters are hard to find. Second, there is a tendency by teachers to write only factual recall questions. Of all the selection-type objective tests, multiple-choice items require the most time for the student to respond. Finally, testwise students perform better than do non-testwise students.

CLASSROOM CASE STUDY

PERFORMANCE-BASED ASSESSMENTS

Nadine Espinoza, a sixth-grade teacher, has been introducing her students to a number of algebraic topics in math class. Recently, the students have been studying single-variable equations.

This year, instead of her usual end-of-unit written test, Ms. Espinoza has decided to substitute a performance-based project. She wants to show her students where the concepts they are learning in class can be found in real-life situations. For their end-of-unit project, the students will be asked to collect information about how a family's cutting energy and fuel costs can help them save for a fantasy vacation. Specifically, students will collect data about their families' electricity and automotive fuel consumption and costs for a month.

Using these data, each student will develop an equation to determine the average amount of electricity and fuel his or her family uses on a daily basis. Each student will decide on a reasonable dollar amount that his or her family can put aside each month toward their

fantasy vacation. The student will then create an equation to show how much needs to be cut from his or her family's daily electricity and fuel consumption to save that amount. Students will be asked to graph the information they have gathered and to incorporate the graphs and data into a brochure that can be used to persuade their families to reduce electricity and fuel consumption to save for the fantasy vacation.

FOCUS QUESTIONS

1. What are the advantages of Ms. Espinoza using a performance-based assessment instead of the traditional paper-and-pencil exam? What disadvantages might exist with a performance-based assessment?
2. Think of some examples of performance assessments that might be used in the following content areas: science, language arts, social studies, and/or foreign language.

Consider the following guidelines when writing multiple-choice questions:

1. Use either a direct question or an incomplete statement as the item stem. A direct question is often preferable.
2. Write items in clear and simple language.
3. Make the response choices as brief as possible.
4. Include only one correct or best answer to every item.
5. With an incomplete statement, place the choices at the end of the statement.
6. Include four to five response choices.
7. State only plausible response choices.
8. List the response choices below the stem in a vertical, easy-to-read format.
9. Use letters for the response choices.
10. Avoid patterns of answers.

Matching. Matching questions consist of a set of directions, a list of stems or numbered items (e.g., statements, phrases, words, incomplete sentences), and a list of choices (e.g., words, phrases, numbers). Students are required to make an association between each premise and the choices.

Matching questions are well suited to test knowledge in terms of definitions, dates, names, events, and other matters involving simple relationships. Because they require relatively little reading time, many questions can be asked in a limited amount of testing time. Matching questions are also easy to score. The range of test material can be broad, and guessing is limited.

There are some disadvantages to matching questions. They may encourage serial memorization rather than association. It can also be difficult to get clusters of questions

that are sufficiently alike that a common set of responses can be used. It is sometimes difficult to avoid giving clues, and too many items may be confusing to the students. Furthermore, matching questions are not well adapted to measure students' understanding of concepts or their ability to organize and apply knowledge.

The following guidelines will assist you in preparing matching questions:

1. List the homogeneous premises on the left and the options to be matched on the right.
2. List the premises and choices in some logical or systematic order (e.g., alphabetically, chronologically).
3. Keep the list of premises and responses relatively short; the optimum is five to eight items to be matched.
4. Provide extra responses to reduce guessing.
5. Be sure the stem includes the longer and more involved statement, and the response choices are short and simple.
6. Number each stem, and use capital letters for each response choice.

True-False. A true-false item is simply a declarative statement to which the student responds by indicating whether it is true or false. One variation asks the student to correct false statements to make them true; these are often called modified true-false.

True-false questions have a number of advantages, including the fact that they are good for students who are poor readers; can cover a large amount of content in a given testing period; generally provide high reliability; can be scored quickly and reliably; are adaptable to most subject areas; and can be constructed to measure the higher mental processes of understanding, application, and interpretation.

There are several disadvantages in the use of true-false questions. Scores may be unduly influenced by guessing; true-false items are often susceptible to ambiguity and misinterpretation, resulting in low reliability; the questions lend themselves to cheating more often than do some other types of questions; and some statements are not entirely true or false, and thus specific determiners are often added to the questions.

To help overcome some of the disadvantages, the following guidelines should be used when writing true-false items:

1. Be certain that the item is absolutely true or false, without any qualifications or exceptions.
2. Avoid loosely worded and ambiguous statements.
3. Highlight the central point of the question by placing it in a prominent position in the statement.
4. Avoid negative statements whenever possible.
5. Avoid making true statements consistently longer than false statements.
6. Avoid a disproportionate number of either true or false statements.

Short Answer. Short-answer or completion questions require students to provide a word or phrase. They are typically in the form of a direct question or a completion statement.

Short-answer questions reduce the chance that students will guess the correct answer, and test items are relatively easy to construct. On the other hand, short-answer items are more difficult to score than other types of objective test items. There is also an emphasis on vocabulary and recall of information.

Use the following guidelines when constructing short-answer questions:

1. Generally, it is better to use a direct question rather than a completion statement.
2. Use short-answer questions to measure only the recall of important information.



VOICES from the Classroom

ELICIA LIGON, middle school science and English teacher, Ponca City, Oklahoma

INCLUDING A VARIETY OF TEST QUESTIONS

When creating a teacher-made test, I try to include questions of all types. There will be short sections of each of the following: multiple choice, true-false, short answer (a sentence or two), an activity (such as a graph or data table), and a short essay question.

I believe this variety helps test all aspects of a student's learning. I also think it puts students at ease to see at least one type of test question that they are comfortable with. These tests take time to prepare, but they seem more like a true measure of student learning in your classroom.

3. Have the statement lead to one or two specific words or phrases.
4. For completion statements, in general, place the blanks near or at the end of the sentence.
5. For completion statements, do not require more than one or two completions to be made in any one test item.
6. When omitting words to make an incomplete statement, leave enough clues so that students know the answer they provide is correct.

ASSEMBLING THE TEST

The mechanical features of a test—how the questions are grouped and arranged, what directions are given, and how the test is formatted—are no less important than the test items themselves. Careful attention to these aspects will enhance the value of the test as an evaluation tool and will also help save you time and effort.

Grouping and Arranging Test Items. For most classroom purposes, test items can be arranged by a systematic consideration of the type of items used, the learning outcomes measured, the difficulty of the items, and the subject matter measured. First, all questions of an item format should be grouped because doing so requires the fewest sets of directions, is the easiest for the students because they retain the same mental set throughout each section, and greatly facilitates scoring. Second, when two or more item formats are used in a test, they should be sequenced in the following order (Miller, Linn, & Gronlund, 2013):

1. True-false or alternative response items
2. Matching items
3. Short-answer items
4. Multiple-choice items
5. Interpretive exercises
6. Essay questions

Arranging the sections of a test in this order produces a sequence that roughly approximates the complexity of the learning outcomes measured, ranging from simple to complex.

Third, within each item format, questions related to the same instructional objective should be grouped. Fourth, items within each item format should be arranged in order of ascending difficulty. Therefore, both the sections and the items within the sections are arranged in an ascending order of difficulty throughout the test.

Writing Test Directions. Directions should indicate what the students need to do, how they are to do it, and where they should record their answers. Furthermore, the directions should indicate the time to be allocated to the test, the value of each test item, and whether students should guess at any answers they are unsure of. When determining test directions, teachers can use the following guidelines:

1. Provide a specific set of written directions for each item format.
2. Indicate how the student is to record the answers.
3. Indicate the point value of each question in an item format.
4. Indicate the purpose of the test, the length of time available, and whether students should guess. These issues can be announced verbally rather than being written on the test paper.

Formatting and Reproducing the Test. The visual display of test items on the paper can affect the time and effort the students expend taking the test and also the teacher's time during scoring. Use the following guidelines when formatting, typing, and reproducing a test:

1. Space items so that they can be easily read, answered, and scored.
2. Number items consecutively throughout the test.
3. Make sure all items have generous borders.
4. Keep all stems and options on one page.
5. For matching questions, be sure the list of premises and choices are on the same page.
6. The most convenient method of response in true-false, matching, and multiple-choice questions is circling the correct answer.
7. Avoid a definite response pattern to the correct answer.
8. If no answer sheet is used, one side of the page should be used for responses to all objective questions, regardless of the item format.
9. Make copies with clear, legible print.

ADMINISTERING THE TEST

The physical and psychological environment should be conducive for students to demonstrate their achievement of learning outcomes. Bearing this in mind, consider the following guidelines when administering a test:

1. Provide comfortable testing conditions, paying attention to such factors as adequate light, ventilation, temperature, noise level, and work space.
2. Do not promote test anxiety by giving warnings about the importance of the test or by threatening students.
3. Avoid giving a test just before or after a long vacation or some other important school event.
4. During the test, provide reminders about the time remaining.
5. Discourage cheating through careful proctoring and other means.
6. Do not talk unnecessarily before the test.
7. Keep interruptions to a minimum during the test.

8. Avoid giving students hints about any test item.
9. At the start of the test, indicate what the students should do once they complete the test.

SCORING THE TEST

Guidelines were presented earlier for evaluating essay tests. Scoring objective tests can be done by machine or by hand. Some schools have facilities to score tests by machines that read students' responses off computer cards.

When scoring tests by hand, you have a few options. If the students write their answers on the test sheets themselves, you can simply create an answer key by using a blank test to write the correct answers. It is extremely helpful, as noted earlier, to have all the answers in one column, such as on the left side of the page. In this way, you can hold the answer key next to the student's test and mark the errors. If the student is asked to write the letter of the correct answer, you should draw a line through the incorrect answer and then add the letter of the correct answer. But if all letter choices are given at the left of the question and students are expected to circle the correct answer, you can easily see an incorrect answer and then indicate the error by circling the correct answer. In this way, in one stroke of the pen, the student knows the choice was incorrect and is also given the correct answer.

A scoring stencil can also be used in a similar way to cover all choices except the correct answer. You then can see which items were correctly marked by the students and place a mark on the correct choice when the students have chosen another response.

You can have students place their answers to objective questions (multiple choice, true-false, matching) on computer cards and use scanning systems to score the tests. This can save considerable time, especially when you have many tests to score, such as at the middle and secondary levels. The cards that the students complete are run through a scanning machine, and each card is scored. Most scanning systems can generate an item analysis of the questions and a class roster with the score. Some scanning systems allow entering the scores from performance-based assignments (essays, special projects, oral exams, homework) with the machine-scored questions to obtain a combined score.

Regardless of the scoring method selected, the scoring key or answer key should be prepared and checked well in advance of the administration of the test. Generally, each item should have equal weight.



Achievement Tests

Standardized tests are those that are prepared by testing agencies with standard items, directions, testing conditions, and scoring procedures. The tests are typically created by test-preparing agencies, such as the Educational Testing Service. The test items are of high technical quality (thoroughly developed, pretested, and selected by the testing agencies). The directions for administering and scoring are so precisely stated that the procedures are standard for different users of the test. Test manuals and other accessory materials are also included as guides for administering and scoring the test, evaluating its technical qualities, and interpreting and using the results (Miller et al., 2013).

Standardized achievement batteries are used in elementary and secondary schools. The term **test battery** in this context means a coordinated series of tests covering different content areas and multiple grade levels (e.g., a series of tests in science in each elementary grade, or a series in mathematics in each grade). Widely used standardized achievement tests include the Iowa Tests of Basic Skills, Stanford Achievement Test Series, Metropolitan Achievement Tests, TerraNova, and California Achievement Tests.

It has been common practice for state departments of education, with the authority of state and federal legislation, to use standardized tests to diagnose and evaluate students' academic progress on a yearly basis. The results of these tests are given to schools

Achievement Tests

What are some standardized tests that you have encountered in your schooling? This video explains how standardized tests (or norm-referenced tests) are constructed, why they are used, and how to prepare for them. Do you think standardized tests are appropriate for the different purposes mentioned in the video?

and teachers. In many states, test scores are summarized by the school and by subgroups within schools. Each school is compared to other schools in the state. These comparisons are sometimes published in newspapers. Test results are often used to make high-stakes decisions about students, teachers, and schools. Thus, standardized achievement tests are often referred to as **high-stakes tests**.

Beginning teachers will be expected to understand and administer standardized tests that have been selected by the district for use. Further, they will be expected to use the test results to improve teaching and learning, and to communicate the results to students and their parents. Some school districts hold teachers accountable for their students' success on these tests.

TYPES OF ACHIEVEMENT TEST SCORES

Since teachers are expected to use the test results, it is important to understand the types of scores that are reported from achievement tests. There are five different types of scores that may be reported on norm-referenced tests (Popham, 2014; Stiggins & Chappuis, 2012).

- **Raw score.** When the students take the test, the number of items that they answer correctly is called the **raw score**.
- **Percent correct.** **Percent correct** refers to the percentage of the test items that the student answered correctly. This score is used to determine mastery over the objectives. Students are judged to have mastered the objectives if they correctly answer a certain percentage of the items covered on the test. The cutoff for mastery on a given test, for example, may be 75 percent of the answers correct.
- **Percentile score.** The **percentile score** (sometimes called the percentile rank) tells us what percentage of the norm group a student with a given raw score outperformed. A student with a percentile score of 56 percent, for example, outscored 56 percent of the other test takers in the test's norm group. Thus, the percentile score gives a straightforward comparison of student-to-student performance.
- **Stanine.** This is a less precise score scale that is based on percentile rank. The percentile scale is divided into nine segments, each of which represents a "standard nine," or abbreviated, a **stanine**. Students scoring with the lowest percentile scores would be in the first stanine (well below average), while students scoring with the highest percentile scores would be in the ninth stanine (well above average).
- **Grade equivalent scores.** The basis of the **grade equivalent score** is the performance of the students in the norm group at specified grade levels. For example, a student with a grade equivalent score of 6.4 in reading is said to have scored about the same as students in the norm group who were in their fourth month of sixth grade.

THE TEACHER'S ROLE IN STANDARDIZED TESTING

Teachers play several important roles concerning standardized testing. These include preparing students for the test, administering the test, and then using test scores in various ways (McMillan, 2014; Miller et al., 2013).

Preparing the Students. A number of actions can be taken before test day to adequately prepare students for the standardized test. Students will perform better if they understand the reasons for the test, know about the test format and characteristics, know how they can prepare, and are prepared to use effective test-taking skills.

- *Discuss the reasons for taking the test and communicate a positive attitude toward standardized tests.* If you convey negative comments to students about the test, then students will also adopt such an attitude and may not try their best. Instead, discuss the reasons for the test and the ways the test results will help improve teaching and learning. Convey an attitude of challenge and opportunity. Emphasize that it is important for students to try to do their best, not just to obtain a high score.

- *Familiarize students with the test format and provide practice opportunities.* Acquaint students with the test format, number of sections and questions, time limits, directions, types of test items, and other factors. Make sure students know what the test covers. Whenever possible, let them practice on previously released tests so that they become familiar with all aspects of the test. Let them see the blank answer sheet that will be used.
- *Talk to students about the test.* Motivate students to put forth their best effort. Reassure students that some test anxiety is normal and can provide energy to help them perform better. Emphasize that a good night's sleep and a healthy breakfast or lunch will help students stay alert and at their best for the test.
- *Help students learn effective test-taking skills.* Most experts agree that students who have good test-taking skills do better on standardized tests than do those students with poor skills (Russell & Airasian, 2012; McMillan, 2014). Many districts have designed programs specifically to improve test-taking skills, and these materials may be available to classroom teachers when working with their students. Some test-taking skills include strategies for using test time, avoiding errors, guessing at answers, and using deductive reasoning (Brookhart & Nitko, 2015).

Administering the Test. Administering a standardized test is more formal than administering a teacher-made test. Standardized tests have very specific instructions for teachers when administering the tests, and the stakes for student performance are higher than for teacher-made tests. For these reasons, it is important to pay attention to the following issues when administering the standardized test (McMillan, 2014; Popham, 2014):

- *Establish a suitable environment for the test.* Just before administering the test, make sure the lighting is adequate, students have acceptable working space, desks are cleared, and distracting objects are removed. Put a "Do Not Disturb" sign on the classroom door when the test is being conducted.
- *Follow the test administration directions precisely.* Standardized tests have been developed with directions designed to standardize the conditions under which students in different classes and schools take the test. Teachers must follow the instructions developed by the publisher. The directions indicate what to say, how to respond to student questions, and what to do as students are working on the test. Read the directions to the students word for word as specified. During the test, teachers may answer questions about the directions or procedures about answering items, but they may not give clues or hints.
- *Time the test accurately.* Most standardized tests have time limits, and it is important to adhere to the time. Mark down the starting time when you say "Go" and also make a note of the ending time. When you say "Stop," make sure students really do stop.
- *Monitor the situation.* After students have started the test, teachers should circulate around the room. Notice if students are working on the correct sections of the test and are recording answers appropriately. Circulating should also discourage cheating. Be prepared for students who finish early because their actions may be a source of distraction for other students who have not yet completed the test.
- *Make notes of any unusual circumstance or event.* While observing students as they take the test, teachers may see some unusual behavior or events that could affect the students' performance, such as construction noise outside or several interruptions at the classroom door. Or individual students may have some difficulty, such as a contact lens causing a vision problem. It is best to record these behaviors or events for use in subsequent interpretation of the results. Interruptions should also be recorded.
- *Take actions after the test has been completed.* Retrieve all test materials once the test has been completed and store them securely until you submit them to the school's central office. File any notes you made during the test administration about any unusual behaviors or events. Submit all materials to the school's central office.



VOICES from the Classroom

PATRICIA SMITH, sixth-grade teacher, Monroe Township, New Jersey

USING STANDARDIZED TEST DATA TO INFORM INSTRUCTION

At the beginning of the school year, the district provides us with student assessment data from standardized testing. The students also take a computerized benchmark test in language arts and math during the first month of school and complete a writing prompt that is scored using a holistic rubric. Classroom teachers are required to keep a spreadsheet of all of the data to be used to inform instruction. With this data collected in one resource for the entire class, it is easy to identify areas by student or class that are in need of instruction.

I identify students who show areas of weakness in language arts and wherever possible match the deficiencies to our curriculum and the state standards. I keep a separate section in a binder where these students' data are highlighted, and I document what I have done to address their needs. Through

differentiated instruction, I can provide these students with reteaching or practice on a skill while other students work on enrichment activities. I keep folder activities for this purpose, and each student has folders assigned for him or her to work on. I can also post assignments online through various programs, and all students will work on laptops but on different assignments. This is also a time for small-group or individualized instruction. I also use the information to group students heterogeneously so that group work supports struggling students.

I constantly update my spreadsheet and binder to reflect the most recent data and mark each student's progress. This information also helps me conference with parents. Having accurate and meaningful data is a valuable tool for informing classroom instruction.

Using the Test Scores. Published tests of achievement and ability can play an important role in the school's evaluation program. The test results also can be useful for individual classroom teachers in the following ways (Brookhart & Nitko, 2015; Miller et al., 2013; Waugh & Gronlund, 2013).

Guiding Instructional Planning. Test results provide a data source for instructional planning and help teachers meet students' learning needs. Take the following actions as you consider standardized test results when planning instruction.

- *Identify the level and range of ability among students in the classroom.* Instructional plans take into account the students' learning abilities and their present levels of achievement. Test results provide evidence for the students' level of achievement and can serve as the basis for future instructional planning.
- *Identify areas of instruction needing greater emphasis.* If the published tests were selected in accordance with the school's curricular objectives, weaknesses will be revealed by those areas of the test in which students perform poorly. Published tests are especially helpful in appraising strengths and weaknesses in learning. Based on the test results, teachers may need to make adjustments in their selection of content to address any weak areas.
- *Identify discrepancies between learning ability and achievement.* Large discrepancies in a student's perceived learning ability and the achievement test score may be a sign that a student is underachieving. However, a number of factors may contribute to a student's low test performance, and teachers may seek additional information before modifying instructional plans.
- *Clarify and select instructional objectives.* The results of published tests provide evidence that can help in the selection of objectives at a particular grade level or in a

particular course. Sixth-grade test results on the Iowa Tests of Basic Skills in science, for example, may provide information that fifth-grade teachers can use when selecting instructional objectives. Low performance in sections of that science test may suggest changes in instructional objectives in fifth grade to provide better preparation, and changes in the school's science curriculum also may be necessary.

Identifying the Needs of Exceptional Children. For exceptional students, published tests can be helpful in identifying problems of learning and development so that special provisions can be made for meeting specific needs. This information can be useful when preparing an individualized educational plan for each student.

Providing Parents with an Independent Source of Information. By means of report cards, parent-teacher conferences, and other means of parental contact, teachers inform parents about the educational progress of their children. Standardized test results are a useful supplement to report card grades and other measures that are reported. They have the added advantage of being independent of teacher opinions.

Key Terms

Anecdotal records	Extended response essay questions	Participation chart	Reliability
Checklist	Formative assessment	Percent correct	Restricted response essay questions
Classroom assessment	Grade equivalent score	Percentile score	Rubric
Criterion-referenced evaluation	High-stakes tests	Performance-based assessment	Standardized tests
Data-driven decision making	Measurement	Portfolio	Stanine
Essay questions	Measurement error	Practicality	Summative assessment
Evaluation	Norm-referenced evaluation	Preassessment	Table of specifications
	Objective questions	Rating scale	Test battery
		Raw score	Validity

Major Concepts

1. Classroom assessment is the collection, evaluation, and use of information to help teachers make decisions that improve student learning.
2. Measurement is the process used to obtain data concerning student learning. Evaluation is the process of making a value judgment about student learning based on the data.
3. Data-driven decision making means that teachers use information from many sources to make decisions about planning, teaching, and assessing.
4. There are three types of assessment: preassessment, formative, and summative. Each serves a different purpose, and each is conducted at different times.
5. Even before beginning instruction, teachers need to establish a framework for classroom assessment.
6. When using performance-based assessment, teachers must specify the criteria by which the student products or performances will be rated, and then they must prepare the actual rating forms.
7. A table of specifications identifies objectives in all three learning domains that students should attain by the end of instruction and also the content related to these objectives.
8. Student assessment should be related to instructional objectives that are included in the table of specifications.
9. Guidelines for planning the classroom test, selecting and preparing test items, assembling the test, and administering and scoring the test can aid in effective measurement of student learning.
10. Standardized tests are those that are prepared by testing agencies with standard items, directions, testing conditions, and scoring procedures. Teachers have responsibilities to prepare students for the tests, administer the tests, and then use the test scores in various ways.

Discussion/Reflective Questions

1. During a class session, what are some ways that teachers could use formative assessment to check for student understanding?
2. Why is it useful to have a variety of assessment measures when evaluating student learning?
3. To what extent should assessments be adjusted for students with special needs? Is there a limit to how many adjustments should be made for a student?
4. What are the advantages and disadvantages of using products and performances to assess student learning?
5. What are the advantages of using a table of specifications when preparing a test?

Suggested Activities

FOR CLINICAL SETTINGS

1. Prepare a table of specifications for this chapter that includes learning objectives and content.
2. For a test on this chapter, prepare three questions for each of the following types of questions: true-false, short answer, multiple choice, and essay.
3. Construct a rubric that could be used in assessing students' understanding of a short story.

FOR FIELD EXPERIENCES

1. Ask one or more teachers to discuss the way they evaluate students. Analyze their responses in relation to the framework for assessment discussed in this chapter.
2. Talk with several teachers about the ways that they conduct product or performance assessments. Ask specifically about rating criteria and rating forms.
3. Obtain one or more teacher-made tests and critique them based on the guidelines discussed in this chapter for planning the test, selecting and preparing test items, and assembling the test.

Further Reading

Butler, S. M., & McMunn, N. D. (2006). *A teacher's guide to classroom assessment: Understanding and using assessment to improve student learning*. San Francisco: Jossey-Bass.

A comprehensive guide that shows step-by-step ways to integrate assessment in the classroom. Considers meeting standards, setting expectations, gathering assessment evidence, making sense of the data, and linking assessment to instruction.

Garéis, C. R., & Grant, L. W. (2008). *Teacher-made assessments: How to connect curriculum, instruction, and student learning*. Larchmont, NY: Eye on Education.

Provides very useful guidelines and examples for various types of test questions. Describes how to prepare tests and how to use them in an assessment system.

Miller, M. D., Linn, R. L., & Gronlund, N. E. (2013). *Measurement and assessment in teaching* (11th ed.). Boston: Pearson.

Provides in-depth coverage of all aspects of measurement and assessment. Many examples and guidelines are included.

Waugh, C. K., & Gronlund, N. E. (2013). *Assessment of student achievement* (10th ed.). Boston: Pearson.

Serves as a practical guide for assessing the intended outcomes of instruction. Emphasis on teacher-made assessments and performance assessments.