

CHAPTER 6: DEVELOPING AND REFINING FORMATIVE ASSESSMENTS

ONLINE APPENDIX

ASSESSMENT TYPES—PROS AND CONS

In the following appendix, we briefly review the various types of assessment formats used in formative assessments and offer general guidelines for their development. For more detailed explanations, the books listed in Chapter 6 are highly recommended.

TRADITIONAL ASSESSMENTS

Traditional assessment formats include true/false, matching, multiple-choice, completion, and essay and constructed-response items. We refer to these as *traditional* because they are the assessment formats most familiar to teachers and the ones with which they have the most experience, both as teachers and as students.

The first three of these formats—true/false, matching, and multiple-choice—are classified as *selected-response* items because they ask students simply to choose the best or most appropriate response from among those provided. Completion, essay, and all the alternative assessment formats are considered *supply* or *constructed-response* formats because they require students to construct or produce their own response.

Some writers discourage teachers from using traditional assessment formats. But while they certainly have their limitations, traditional formats are highly appropriate in many instances and serve a variety of useful purposes. When selected-response formats are well constructed, teachers can use them to assess not only students' mastery of content knowledge but also their ability to use that knowledge to reason and solve complex problems. The key is to ensure that the assessment format matches the concepts and skill level addressed.

True/False Items

True/false items are declarative statements that students must judge to be either true or false.

For example:

DIRECTIONS: Circle the T or F in front of each of the following statements depending on whether you believe it to be true or false.

T F 1. Shakespeare's *The Tragedy of Hamlet* is set in Denmark.

T F 2. *The Rime of the Ancient Mariner* is the story of a huge white whale written by Herman Melville.

True/false items are relatively easy to write and are easily scored. One exception is when students are asked to respond by printing either a *T* or an *F*. Invariably, some students will offer the following response, hoping the teacher reads it quickly and judges it to be whatever response is correct:

F

To solve this problem, simply print a *T* and *F* before the statement and have students circle their response, as shown in the earlier example.

Although easy to develop and score, true/false items do have their limitations. First, they typically measure only a small bit of content and are generally appropriate only for lower-level cognitive skills. It's difficult to develop complex, high-level true/false items. A second and perhaps more serious limitation is that students have a fifty-fifty chance of guessing the correct response without any knowledge or understanding of the content. Furthermore, knowing that a particular statement is incorrect offers no guarantee students know what is truly correct. Sometimes teachers try to get around this by asking students to reword those statements they determine to be false in order to make them true. But most assessment experts discourage this practice because it results in extremely complicated scoring procedures. For example, how would you score the responses of students who recognize the statement is false, but the way they revise it is also incorrect?

The best true/false items contain a single, significant idea and are worded precisely so that they can be judged unequivocally true or false. Good true/false items also avoid extraneous clues to the answer—particularly words such as *always*, *never*, *usually*, or *sometimes*—since these often allow students to select the correct response without truly understanding the concept. It's also best to avoid using double negatives since they can make statements difficult to interpret. For example, the statement “You should not teach children to never cross the street before looking both ways” should be judged false, but it's certainly not obvious.

Matching Items

Matching items typically consist of a series of items or premises listed in one column and a series of responses in another. In answering, students match each item with one or, in some cases, a number of the responses. For example:

DIRECTIONS: Match each major act in the left-hand column with the president who signed that act listed in the right-hand column by placing the appropriate letter in the blank before the act. A president may be matched with several or none of the acts.

- | | |
|--|--------------------------|
| ___ 1. Pure Food and Drug Act (1906) | a. Theodore Roosevelt |
| ___ 2. Social Security Act (1935) | b. Woodrow Wilson |
| ___ 3. Federal Highway Act (1956) | c. Franklin D. Roosevelt |
| ___ 4. National Defense Education Act (1958) | d. Harry S. Truman |
| ___ 5. Nuclear Test Ban Treaty (1963) | e. Dwight D. Eisenhower |
| ___ 6. Civil Rights Act (1964) | f. John F. Kennedy |
| ___ 7. Voting Rights Act (1965) | g. Lyndon B. Johnson |
| ___ 8. Medicare and Medicaid Act (1965) | h. Richard M. Nixon |
| ___ 9. Clean Air Act (1970) | i. Ronald Reagan |
| ___ 10. Economic Recovery Tax Act (1981) | j. George H. W. Bush |

Teachers generally find matching items easy to write and score. Matching items are particularly appropriate with vocabulary lists and often can be used to cover a wider scope of material than true/false items. But like true/false items, matching items work best when measuring lower-level cognitive skills and are appropriate only in specific instances—that is, with lists of terms or facts.

The best matching items include homogeneous material so that all responses are likely alternatives. For example, if only one premise asks for a date, and only one date is included among the responses, it doesn't require much skill to determine which response is correct. The lists of items also should be kept fairly short, with the brief responses placed on the right-hand side of the page. Because students typically work from left to right on a page, formatting matching items in this way significantly reduces the time required for the assessment.

The best matching items also include either more or fewer responses than premises. If there is a one-to-one correspondence between premises and responses, students will be able to determine the correct answers to some items simply through a process of elimination. In addition, directions should be provided that specify the basis for matching and indicate that some responses may be used more than once or not at all.

Tips for Good Matching Items

1. Construct responses to be fairly homogeneous so that all responses are likely alternatives.
2. List long premises on the left and short responses on the right to reduce students' response time.
3. Include either a larger or smaller number of premises as responses.
4. Provide specific directions for the matching process.

Multiple-Choice Items

Multiple-choice items consist of a stem that presents a problem situation and several alternatives that provide possible solutions to the problem. The stem is generally a question or an incomplete statement. The alternatives include one correct answer to the question and several other plausible wrong answers referred to as “distracters.” For example:

DIRECTIONS: For each of the questions that follow, select the best answer. Place the letter of your answer on the blank in front of the item.

- ___ 1. Using semantic differential techniques, it was discovered that *good* has slightly male overtones, while *nice* has slightly female ones. This is an example of what kind of word meaning?
- a. Explicit
 - b. Symbolic
 - c. Conceptual
 - d. Denotative
 - e. Connotative
- ___ 2. When children first learn a new word, they tend to overextend its use. In learning theory terms, this is very much like _____.
- a. generalization
 - b. discrimination
 - c. operant learning
 - d. latent learning
 - e. cognitive learning

Multiple-choice items have several advantages over other assessment formats. First, they can be constructed so that a single item measures more than one concept. When students answer a multiple-choice item, they must determine not only which alternative is correct but also that the others are incorrect. Hence a single item can cover a broad range of material. Second, carefully constructed

multiple-choice items can assess higher-level cognitive skills, especially transfer and application skills. For example, consider the following item:

- ____ 1. A researcher found a correlation of $+0.68$ between students' scores on a test of creativity and measures of their verbal fluency. Which of the following would be a valid conclusion to draw?
- More creative students generally have higher verbal skills.
 - As students become older, their verbal skills increase.
 - Increasing students' verbal skills will help them become more creative.
 - Approximately 68 percent of creative students have high verbal skills.
 - Fostering students' skills in art or music will likely enhance their verbal fluency.

All of these responses may be plausible. To answer this item correctly, however, students must recognize (1) age was not considered in the description of the study (eliminating *b*), (2) correlation does not imply causation (eliminating *c* and *e*), and (3) correlation is unrelated to percentage (eliminating *d*).

A third advantage of multiple-choice items is that unlike supply or constructed-response items, they are quick and easy to score. Computers with optical scanning technology can score multiple-choice items and provide detailed item analyses in a matter of seconds. And finally, multiple-choice items allow educators to gain the most amount of information about students' performance on the skills they measure in the least amount of time. These traits make multiple-choice items an exceptionally efficient form of measurement and the reason they are used as a major part of nearly every standardized achievement assessment.

From an instructional perspective, multiple-choice items have one other important advantage: They can be constructed to provide important *diagnostic information*. For example, consider the following mathematics item:

1. $2.3 + 0.15 = \underline{\hspace{2cm}}$.

- 3.8
- 0.38
- 1.73
- 0.245
- 2.45

The most common error students make in answering a mathematics problem like this is not aligning the decimal points before adding. Thus, if a student selects alternative *a* or *b*, the teacher can be fairly certain of the error that was made

and can offer specific corrective help. Students who select alternative *c* or *d* may have similar but clearly more complex misunderstandings.

Advantages of Multiple-Choice Items

1. Multiple concepts can be assessed in a single multiple-choice item.
2. Carefully constructed multiple-choice items can assess higher-level cognitive skills.
3. Multiple-choice items can be quickly and easily scored.
4. Multiple-choice items allow the most information about students' performance to be gathered in the least amount of time.
5. Well-constructed multiple-choice items can be used to diagnose students' learning problems.

Among the assessment formats, however, good multiple-choice items are probably the most challenging to construct. It can be difficult to develop clear and concise item stems as well as alternatives to the correct response that are plausible yet not unfairly misleading. This problem frequently leads to creating more than one correct response. The following books offer helpful guidelines for preparing good multiple-choice items:

Bloom, B. S., Madaus, G. F., & Hastings, J. T. (1981). *Evaluation to improve learning*. McGraw-Hill.

Lane, S., Raymond, M. R., & Haladyna, T. M. (2016). *Handbook of test development* (2nd ed.). Routledge.

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

Additional recommendations are available from these online resources:

Brame, C. (2013). *Writing good multiple-choice test questions*. Vanderbilt University Center for Teaching. <https://cft.vanderbilt.edu/guides-sub-pages/writing-good-multiple-choice-test-questions/>

Centre for Teaching Excellence. (n.d.). *Designing multiple-choice questions*. University of Waterloo. <https://uwaterloo.ca/centre-for-teaching-excellence/teaching-resources/teaching-tips/developing-assignments/assignment-design/designing-multiple-choice-questions>

Malamed, C. (2010). *10 rules for writing multiple-choice questions*. The eLearning Coach. http://theelearningcoach.com/elearning_design/rules-for-multiple-choice-questions/

Among the many suggestions provided by the authors of these books and online resources, some of the most important include the following:

1. *The stem of the item should present a single, clearly formulated problem.* It's best to state the stem of the item so clearly that students understand the problem without looking at the alternatives. This clarifies the task for students and enhances the validity of the item.
2. *As much of the wording as possible should be placed in the stem.* Putting most of the wording in the stem serves to further clarify the problem, avoids the repetition of material, and reduces the time students need to read the alternatives.
3. *The stem should be stated in positive form whenever possible.* Generally, positively phrased items measure more important learning outcomes than negatively stated items. Furthermore, being able to identify responses that do not apply provides no assurance that students possess the desired understanding. When negative wording is used, it should always be emphasized with underlining, *italics*, **boldface**, or UPPERCASE lettering.
4. *All alternatives should be grammatically consistent with the stem and parallel in form.* Inconsistencies in tense, articles, or grammatical form among the alternatives can provide clues to the correct response, or at least make some of the incorrect alternatives less effective.
5. *Verbal clues that might enable students to select the correct response or eliminate incorrect alternatives should be avoided.* The wording of an item often provides clues to the correct response. Some of the most common verbal clues include:
 - a. Similarity of wording in both the stem and the correct response.
 - b. Stating the correct response in textbook language or stereotyped phraseology.
 - c. Stating the correct response in greater detail (the longest alternative is likely to be the correct one).
 - d. Including absolute terms in incorrect alternatives (e.g., *all*, *only*, and *so on*).
 - e. Including two alternatives that have the same meaning.
6. *Incorrect alternatives should be plausible and attractive.* All alternatives should be plausible, similar in length, and similar in complexity. In addition, including common misconceptions or errors among the alternatives often helps identify the students' specific learning difficulties.
7. *"All of the above" should be avoided, and "None of the above" should be used with extreme caution.* Using "All of the above" as an alternative makes it possible for students to answer the item on the basis of partial information. Students can detect that "All of the above" is the correct response simply by noting that two of the alternatives are correct, or that it is an incorrect response by recognizing that one of

the alternatives is incorrect. Using “None of the above” may measure students’ ability to recognize incorrect responses rather than their understanding of the concept.

8. *The position of the correct response should be varied in a random manner.* The correct response should appear in each alternative position about the same number of times, but its placement should not follow any particular pattern.
9. *Each item should be independent of the other items in the assessment.* The information given in the stem of one item shouldn’t help students answer another item. It’s also best to avoid item chains where the correct response to one item depends on knowing the correct response to a preceding item. Chains of interlocking items overly penalize students who are unable to answer an earlier item in the chain.

Although it’s challenging to develop good multiple-choice items, their advantages of broad scope and objective scoring make them ideal for assessing important concepts and skills in many subjects. In addition, because most standardized achievement assessments require students to respond to multiple-choice items, practice with these types of items on classroom assessments helps students prepare for those experiences.

Completion Items

Measurement experts describe two general types of completion items. The first consists of questions or incomplete statements for which students are to provide the appropriate words, numbers, or symbols. The second involves problems, especially in mathematics, where students are expected to respond with a complete answer and sometimes to show their work in finding the solution. The following are examples of both types of completion items:

DIRECTIONS: Complete each of the following statements by writing the correct word or phrase in the blank. Be sure to use correct spelling.

1. The largest group under which organisms are classified is the _____.
2. All of the plants in an area are referred to as the _____.

Or, as is common in mathematics:

DIRECTIONS: Solve each problem and record your answer in the space provided. Be sure to show your work and, in the case of fractions, be sure to reduce your answer to the simplest form.

$$\begin{array}{r} 1. \quad 53 \\ \quad - 17 \\ \hline \end{array}$$

$$2. \quad \frac{1}{2} + \frac{2}{3} = \underline{\hspace{2cm}}$$

Completion items are fairly easy to write and sometimes can be used to assess higher-level cognitive skills. Specifically, they require students to recall correct information rather than simply to recognize and select it. However, they can present scoring complications.

One potential scoring problem occurs when there is more than one correct response, particularly if synonyms are available. Other problems arise over spelling errors and the interpretation of unclear or illegibly written responses. In mathematics, teachers face the added dilemma of how to allocate partial credit, especially when the solution to a problem involves multiple steps. Because of these potential difficulties, teachers generally restrict their use of completion items to situations where the learning goals or standards require students to remember or recall information, where computation problems involve multiple steps and each step is important, or where a selection type of item would make the answer obvious.

The best completion items are stated so that only a single, brief response is possible. The words supplied by students should relate to the main point of the statement and should be placed at the end of the statement. Extraneous clues to the answer, such as the use of *a* or *an* and singular versus plural verbs, should be avoided. In addition, directions should be clear and explicit so that students know exactly what is expected in their response.

Essay and Constructed-Response Items

Essay and constructed-response items ask questions or pose problems that require students to produce their own written responses. Sometimes these are referred to as “open-ended response” or “free response” items. In most cases, students are free to decide how to approach the question or problem, what information to use, and how to organize their response. For example:

DIRECTIONS: Write an essay comparing the struggle for independence in the United States in the eighteenth century and in Vietnam in the twentieth century. Describe and give examples of at least three ways in which these struggles were similar and three ways they were different. Your response will be scored in terms of its completeness, the appropriateness of the examples, and how well it is organized. (30 points)

Or, in mathematics:

DIRECTIONS: Answer the following problem. Be sure to show your work in each step and label your answer.

Monica buys two apples at lunch that cost \$0.65 each. She gives the cashier \$2.00 to pay for the apples. Calculate how much change Monica should receive and list a combination of bills and/or coins that would make that amount. (15 points)

Well-written essay and constructed-response items can be used to measure a variety of higher-level cognitive skills. Students' abilities to analyze situations or synthesize different aspects of their learning are difficult to assess with selection types of items. The open response format of essay and constructed-response items, however, places a premium on students' abilities to produce, integrate, organize, and express ideas and concepts in their own words.

At the same time, essay and constructed-response items also have their drawbacks. One problem is that they provide only a limited sample of students' learning. Because students require more time to respond to essay and constructed-response items, typically only a small number of these questions can be included in an assessment. This restricts the learning that can be assessed to relatively few areas.

Another shortcoming is that students' responses to essays and constructed-response items can be distorted by differences in students' writing abilities. Since students must state responses to constructed-response questions in their own words, poor expression and errors in punctuation, spelling, and grammar often lower some students' scores. This is especially true for students who may be English learners. On the other hand, students with excellent writing skills who can express themselves well may be able to bluff their way through questions and inflate the scores they receive.

A third related shortcoming of essay and constructed-response items is the potential influence of subjectivity in scoring. Besides the responses being difficult and time-consuming to score, consistency in scoring can be difficult to achieve. As the style and content of responses shift from paper to paper, teachers' grading standards tend to shift as well (Clouser, 2000). Even differences in students' names and the neatness of students' handwriting can affect scores (Quinn, 2020a; Sprouse & Webb, 1994; Sweedler-Brown, 1992). Later in this appendix, we will discuss ways to limit the influence of such factors through the use of well-designed rubrics.

Despite these shortcomings, essay and constructed-response items are an effective way to assess many kinds of learning. The best essay and constructed-response items are designed to measure *only* complex learning and are *not* used to assess students' recall of terms or knowledge of specific facts or principles. Other item types do this much more effectively and efficiently. Essay and constructed-response items should be restricted to assessing the higher-level cognitive skills associated with transfer, application, analysis, and synthesis.

The questions asked or problems posed in essay and constructed-response items should present a clear and precise task. Students also need to be provided with ample time to respond to each question or problem. This is particularly important since essay and constructed-response items require time for students to think about, organize, and then compose their responses.

Scoring Criteria for Essay and Constructed-Response Items

The most crucial aspect of all essay and constructed-response items is the scoring criteria. These criteria must be clear, concise, and easy for students to understand. Furthermore, teachers must specify these criteria in advance and communicate them to students as part of the instructional process. In this way, students know precisely how their performance will be judged. This generally requires a detailed scoring “rubric.”

A rubric identifies simply and precisely the important dimensions of an appropriate response and how the responses will be judged. By making the scoring criteria clear, rubrics improve the objectivity of scoring and enhance the clarity of feedback to students on how their responses can be improved.

In her book *How to Create and Use Rubrics for Formative Assessment and Grading*, Susan Brookhart (2013) offers excellent practical advice on the development and use of rubrics. She stresses that effective rubrics include two parts: (1) a coherent *set of criteria* for a response, performance, or piece of work that specify what counts; and (2) *descriptions of levels of quality* for each of those criteria ranging from “Excellent” to “Poor.” According to Brookhart, the key is always to remember that rubrics don’t *evaluate* students’ performance; rather, they *describe* students’ performance.

Essential Characteristics of Effective Rubrics

1. A coherent set of criteria that specify what counts
2. Graduated levels of quality for each of those criteria (Brookhart, 2013)

The scoring criteria for essay and constructed-response items are determined primarily by the learning goals or standards they measure. Hence, the table of specifications (see Chapter 4) becomes the primary guide in establishing these criteria. Many teachers begin this process by developing a model response that clearly meets the learning goals they want to assess. They then explain to students how the model response merits an excellent rating on each of the criteria described in the rubric.

Other teachers introduce the rubric as part of their instruction for the unit. They describe the criteria for judging the overall quality of a response along with the three or four levels of quality for each criterion. They then read a sample response with their students, and together they assign a quality level on each criterion. This is sometimes referred to as the *rating scale approach* (Arter & McTighe, 2001; Stiggins, 1994). Teachers generally find this approach highly engaging for students. Plus, because it guides students in diagnosing learning problems and then prescribing ways to make improvements (Gronlund, 1993), many mastery learning teachers find it particularly effective. An example of a rating scale scoring criterion appears as follows.

Example of a Rating Scale Scoring Criterion	
Score	Description
4	The response is clear, accurate, and to the point. It includes relevant information and offers good support. Important connections are made, and insights important to the objective are included.
3	The response is clear and somewhat focused, but not particularly strong. Only limited support is offered for the points made, and the connections are remote. Few important insights are evident.
2	The response is mostly clear but somewhat confused. The support offered is weak and the connections vague. The insights offered are weak.
1	The response is off the mark. It does not show mastery of the important concepts or contains inaccurate information. No important connections are made, support is missing, and/or no important insights are included.

When scoring essay and constructed-response items, it is usually best to mark all students' responses to one item before proceeding to the next item. This helps maintain consistency in scoring students' responses. Most teachers also find it best to evaluate *content* and *clarity* (mastery of important concepts and skills) separately from *form* (grammar, punctuation, and writing proficiency) since different criteria will be relevant for each.

Whenever possible, it's also best to score responses to essay and constructed-response items anonymously—that is, without knowing the identity of the students. Although many teachers know their students by their handwriting, particularly at the elementary level, anonymous scoring helps reduce unconscious subjectivity and bias (see Quinn, 2020a, 2020b). Having students put their names on the back of papers or on a separate, attached sheet helps reduce such subjectivity. Additional suggestions for the development, scoring, and appropriate use of essay items are available in the books mentioned in Chapter 6, and also in the following:

Arter, J., & McTighe, J. (2001). *Scoring rubrics in the classroom: Using performance criteria for assessing and improving student performance*. Corwin.

Moskal, B. M. (2000) Scoring rubrics: What, when and how? *Practical Assessment, Research, and Evaluation*, 7(3), 1-5.

Panadero, E., & Jonsson, A. (2013). The use of scoring rubrics for formative assessment purposes revisited: A review. *Educational Research Review*, 9(1), 129-144.

Thorndike, R. M., & Thorndike-Christ, T. M. (2010). *Measurement and evaluation in psychology and education* (8th ed.). Pearson.

ALTERNATIVE ASSESSMENTS

Alternative assessment formats encompass a wide variety of assessment methodologies. Collectively, they are referred to as “performance” or “authentic” assessments because they are considered valuable activities in themselves and typically involve the performance of tasks directly related to real-world problems (Linn et al., 1991; Palm, 2008). Alternative assessments provide students with opportunities to display their understanding in thought-demanding ways (Perkins, 1998), and typically engage students in activities that require the demonstration of important skills or the creation of specified products. As such, they are used primarily to assess deeper learning (Harris et al., 2019) and higher-level, complex learning goals that cannot be adequately addressed through more traditional assessment formats.

Alternative assessments are highly appropriate and widely used as formative assessments. Like completion and essay items, they are considered a supply or constructed-response format since all forms require students to construct or produce their response. Like constructed-response items, however, they should be used only to measure higher-level cognitive skills. In other words, alternative assessments should be used primarily to assess learning goals or standards that involve applications, transfer, analyses, and syntheses. Although various forms of alternative assessments can be used to measure knowledge-level learning goals, traditional formats generally do so much more reliably and efficiently.

Alternative assessments include skill demonstrations, oral presentations, task performances/complex problems, compositions or writing samples, laboratory experiments, projects or reports, portfolios, and group tasks or activities. These formats require teachers either to observe students while they are performing or to examine the products students create in order to judge the level of proficiency. Mastery is determined by comparing each student’s performance or product to preset criteria.

Scoring Criteria for Alternative Assessments

The most challenging aspect of the use of alternative assessments, regardless of the form, is devising sound scoring criteria. To guide the development and clarification of appropriate criteria, Jan Chappuis and Rick Stiggins (2017) outline a six-step process:

Step 1. Begin by reflecting on the meaning of excellence in the performance area you have in mind. Tap the professional literature, texts, and curriculum materials for insights. The wisdom of colleagues can be a valuable resource. Also include students as partners in the process. Start by developing a list of key elements and then refine the list as your thinking clarifies.

Step 2. Categorize the list of elements to reflect your priorities. Keep the list as short as possible while still capturing the essence of performance.

Step 3. Define each key dimension in clear, simple language.

Step 4. Find examples of actual performances or products to study. It's best if these examples include contrasting cases—for example, a smooth and accurate jump shot in basketball and a poor one; an outstanding term paper and a very weak one; a student who contributes effectively in a group and one who disrupts the exchange of ideas.

Step 5. Spell out each element in words or illustrations along a continuum of performance using clear language and examples. This will help define the important dimensions of learning to be assessed.

Step 6. Try the performance criteria to see if they capture the essence of the performance or product. Make any adjustments needed to ensure they precisely communicate the dimensions of success. Adjustments can continue if needed during instruction.

These steps show that the process of devising sound scoring criteria is complex and ongoing. While engaging in this process, it's important to keep two things in mind. First, recall Brookhart's (2013) admonition that rubrics *describe* performance rather than evaluate performance. And second, remember the underlying purpose is to help students become better performers and more successful learners. Regardless of the format, formative assessments are learning tools, designed to identify learning successes and point out where additional study and effort are needed.

Skill Demonstrations

In skill demonstrations, teachers typically observe students as they display a particular skill and then offer feedback to students based on their proficiency. Most teachers think of the use of skill demonstrations first in art, music, drama, and physical education. But they are also common in kindergarten and early elementary grades when students have limited writing skills.

Some skills are simple to judge and allow teachers to use a “checklist approach” indicating whether or not students have mastered the skill. This is especially common with basic skills taught in the early elementary grades. A kindergarten teacher, for example, might watch students as they sort objects according to their shape and then draw inferences about students' abilities to recognize distinguishing features and make classifications.

Other skills, however, are more complex and cannot be judged as simply. For example, a physical education teacher might observe students' proficiency in dribbling a ball around a series of obstacles, or a music teacher might watch students play a particular composition. With more complex skills such as these, most teachers use a “rating scale approach” to indicate students' degree of mastery. After observing students display the skill, they judge proficiency on

a continuum of achievement levels ranging from low to very high (Chappuis & Stiggins, 2017). In either case, clear scoring criteria are essential so that students who do not perform the skill at a high level of proficiency can be offered specific guidance on how to improve.

Skill demonstrations are an appropriate assessment format at all levels of education. They offer teachers accurate evidence on students' competence and proficiency on a variety of important learning goals and standards. Furthermore, teachers can combine skill demonstrations with more traditional assessment formats within the same formative assessment to tap a broader range of learning goals. For example, a science teacher may want to assess students' knowledge of a particular laboratory technique, as well their ability to use the technique appropriately and efficiently. To do so, the teacher would divide the formative assessment into two parts. The first part would be composed of matching and completion items that assess students' knowledge of the technique and its purpose. The second part would involve having students actually use the technique in conducting a scientific experiment.

As with essay items, the most challenging aspect of skill demonstrations is developing the scoring criteria. These should be clear, concise, and communicated to students in advance of the assessment. Many teachers actually distribute printed copies of the scoring criteria to students when they teach a skill so that students know better what is expected and how to prepare for the assessment. Following the assessment, students receive feedback based on the criteria indicating the teacher's judgment of their proficiency and areas where improvement may be needed.

The major drawback of skill demonstrations is that in most cases, they must be individually administered. As such, teachers need to plan alternative learning activities for students not engaged in the assessment so they are not standing around, waiting for their turn, and losing valuable learning time. Many teachers avoid this by coordinating assessment tasks with group activities or by making their observations in an unobtrusive way while students are engaged in skill-related work.

Oral Presentations

Teachers sometimes think of oral presentations only as formal addresses or reports delivered to an audience. But oral presentations offer an effective assessment format in many subjects and at all grade levels. Examples include "Show and Tell" activities in a kindergarten class as well as "Sound It Out" activities and other word recognition tasks in elementary language arts classes. Oral reading is a form of oral presentation. Foreign language teachers typically use oral activities in every instructional unit, as do speech teachers. Of course, oral reports, debates, explanations, and organized class discussions are all forms of oral presentation as well.

Oral presentations require teachers to listen to and interpret students' responses, evaluate quality, and then draw inferences about the level of achievement. An added benefit of oral presentations is that they allow teachers to ask follow-up questions or prompt students to expand on their response. Furthermore, oral presentations can be valuable learning experiences for the presenter as well as the audience.

Like other alternative assessment formats, using oral presentations as formative assessments requires clear scoring criteria, prepared in advance and communicated to students. Many teachers combine oral presentations with other traditional assessment formats within a single formative assessment, much like we described with skill demonstrations. This is a frequent practice in foreign language instruction.

Most teachers provide students with immediate feedback on their oral presentations in the form of oral comments or brief discussions on points for refinement. Such immediate feedback can be helpful to students so long as it is constructive and positive in tone, especially if offered publicly. Written feedback from the teacher and/or fellow students is important as well so that students can monitor and direct their improvement efforts.

Michael Russell and Peter Airasian (2011) offer the following as an example of performance criteria for assessing students' oral reports. Note that in this case, the teacher divided the general performance into three areas for physical, vocal, and verbal expression, and then identified specific criteria within each area.

Example of Performance Criteria for Oral Reports

Struggling 1	Developing 2	Proficient 3	Exemplary 4
-----------------	-----------------	-----------------	----------------

1. Presentation

- ___ a. Stands tall and faces the audience
- ___ b. Maintains eye contact with the audience
- ___ c. Uses facial expression and tone to match the report

2. Articulation

- ___ a. Speaks with a clear voice
- ___ b. Speaks loudly enough for the audience to hear
- ___ c. Enunciates words correctly
- ___ d. Varies tone to emphasize points

3. Verbal Expression

- ___ a. Organizes information logically
- ___ b. Chooses precise words to convey meaning
- ___ c. Avoids unnecessary repetition
- ___ d. Uses sentences with complete thoughts or ideas
- ___ e. Summarizes main points at conclusion

Task Performances and Complex Problems

Task performances and complex problems frame a challenge for students and then set the conditions for students to meet that challenge. In mathematics, for example, students may be presented with a structured problem where they must analyze the information provided and follow an appropriate sequence of steps to solve the problem. In other cases, the tasks may be less structured or designed to present students with novel problems for which there may be more than one correct solution.

An example of a task performance/complex problem designed for elementary mathematics is provided as follows:

Example of a Mathematics Problem-Solving Task

A friend shows you the following solution to a mathematics problem:

$$\frac{2}{3} + \frac{3}{4} = \frac{5}{7}$$

Explain to your friend if the answer is right or wrong. Then show your friend how you would think about the problem using an illustration. Draw your illustration in the space provided.

Like the other alternative assessment formats, task performances and complex problems are particularly useful in assessing the more advanced cognitive skills of transfer, application, analysis, and synthesis. Many tasks require students

not only to generate a solution to the problem but also to explain the process they followed to reach their solution. Task performances and complex problems can be combined with traditional assessment formats within a single formative assessment to tap a broad range of learning goals or standards. They are also appropriate at all grade levels.

Task performances and complex problems also have their shortcomings, however. In most cases, they are fairly time-consuming for students to complete and provide only a limited sample of students' problem-solving skills. Some researchers suggest, for example, that approximately ten to fifteen tasks may be needed to get a reliable estimate of an individual student's capabilities in a particular subject area (Shavelson et al., 1991, 1992). Furthermore, task performance and complex problems present unique scoring challenges, especially when there may be multiple correct or appropriate responses.

The guidelines we discussed for developing and evaluating skill demonstrations and oral presentations apply to task performances and complex problems as well. Teachers must describe clear and concise scoring criteria that students understand and can use to guide their work. An example of a rating scale for complex mathematics problems such as the one illustrated earlier is outlined as follows:

Example of an Analytic Rating Scale for Complex Problems

1. Understands the problem
 - 4 - Complete understanding of the problem
 - 3 - Misinterprets a minor part of the problem
 - 2 - Misinterprets a major part of the problem
 - 1 - Completely misinterprets the problem

2. Solves the problem
 - 4 - Describes a correct solution with no arithmetic errors
 - 3 - Generally correct procedure with minor omission or procedural error
 - 2 - Partially correct procedure but with major fault
 - 1 - Totally inappropriate procedure

3. Answers the problem
 - 4 - Provides a correct solution
 - 3 - Offers a mostly correct solution but with minor computational errors or incorrect label
 - 2 - Partially correct solution or partial answer with computational errors or incorrect label
 - 1 - No answer or wrong answer based on an inappropriate plan

Source: Adapted from Szetela & Nicol (1992, p. 42).

Compositions and Writing Samples

Compositions and writing samples are undoubtedly the most common alternative assessment format used in classrooms today. Teachers introduce writing samples in early elementary grades and continue to use them in classes through graduate and professional school. Any time teachers ask their students to develop a brief written report on a particular topic or to compose an original composition about a subject or theme, they are using writing samples to assess students' learning.

Compositions and writing samples share many of the same advantages of other alternative assessment formats. Teachers find them particularly useful in gauging students' abilities to analyze and synthesize what they know about a particular topic. In language arts, compositions and writing samples can be used to tap students' skills in language usage, their ability to present ideas in an organized and coherent way, and their extended use of appropriate detail or evidence in support of a particular point of view. Furthermore, research indicates that writing about content reliably enhances learning in science, social studies, and mathematics across *all* grade levels (Graham et al., 2020). Written communication skills are also one of the most desired attributes of new entrants to the twenty-first-century workforce (Casner-Lotto & Barrington, 2006; Rios et al., 2020).

In addition, compositions and writing samples represent an excellent means of introducing students to the use of technology. Word processing software programs such as Microsoft Word, WordPerfect, AppleWorks, and OpenOffice remove many of the laborious mechanical tasks that traditionally were part of the writing process. Not having to worry about line spacing, margins, and pagination frees students to concentrate on the substance and quality of their writing. Revisions are also easy to make and become a natural part of the writing process.

The procedures for developing and scoring compositions and writing samples are quite similar to those described earlier in this appendix for essay and other constructed-response items. At the same time, the importance of organization and coherence in these longer writing samples usually necessitates additional scoring criteria. These criteria may be *holistic*, involving judgment of the overall quality of the work, or *analytic*, in which various aspects of the work are evaluated separately. Most mastery learning teachers favor analytic criteria because they are more prescriptive and offer students more detailed feedback to guide revisions. The following is an example of a rating scale based on analytic criteria that include clarity and coherence, support, organization, and language conventions.

Of course, compositions and writing samples also have their drawbacks. First, they are especially challenging to score objectively and reliably. Like essay and other constructed-response items, numerous factors can influence teachers' judgment of students' compositions and writing samples. Although clear

Example of an Analytic Rating Scale for Student Compositions

KAS Narrative Rubric—Grade 7

Guiding Principle C3: Students will compose narratives to develop real or imagined experiences or events, using effective technique, well-chosen details and well-structured event sequences.

Scoring Elements	Novice	Apprentice	Proficient	Distinguished
Clarity and Coherence	<p>Minimally establishes a context for the narrative, but may lack focus or be unclear.</p> <p>Makes a minimal attempt to establish a point of view and/or may abruptly or frequently shift focus in narration. Makes a minimal attempt or makes no attempt at engaging reader in the narrative.</p> <p>Misses many or all demands of the prompt/task.</p>	<p>Attempts to establish and maintain a clear and coherent context for the narrative, but may have some lapses.</p> <p>Attempts to establish and maintain a clear and coherent point of view, but may have some lapses.</p> <p>Attempts to engage the reader in the narrative.</p> <p>Attempts to address some demands of the prompt/task.</p>	<p>Establishes and maintains a clear and coherent context for the narrative.</p> <p>Establishes and maintains a clear and coherent point of view.</p> <p>Effectively engages the reader throughout the narrative.</p> <p>Addresses all demands of the prompt/task.</p>	<p>Thoroughly establishes and maintains a clear, credible and coherent context for the narrative. Context is authentic to the genre.</p> <p>Skillfully establishes and maintains a clear and coherent point of view that enhances the narrative.</p> <p>Thoroughly and consistently engages the reader throughout the narrative.</p> <p>Thoroughly addresses all demands of the prompt/task.</p>
Support	<p>Includes minimal or no narrative techniques such as dialogue, pacing and description to develop experiences, events and/or characters. Provides incomplete, inaccurate and/or irrelevant details which do not capture the action and convey experiences and events.</p>	<p>Attempts to use narrative techniques such as dialogue, pacing and description to develop experiences, events and/or characters.</p> <p>Attempts to use relevant, descriptive details to capture the action and convey experiences and events.</p>	<p>Effectively uses narrative techniques such as dialogue, pacing and description to develop experiences, events and/or characters.</p> <p>Effectively uses relevant, descriptive details to capture the action and convey experiences and events.</p>	<p>Skillfully uses narrative techniques such as dialogue, pacing and description to develop experiences, events and/or characters to enrich the narrative.</p> <p>Skillfully uses relevant, descriptive details to capture the action and convey experiences and events.</p>

<p>Organization</p>	<p>Makes a minimal attempt or makes no attempt to introduce a narrator and/or character(s). Ineffectively organizes an event sequence. Ineffectively uses transition words or uses no transition words to convey sequence and signal shifts from one time frame or setting to another. Provides a weak conclusion or lacks a conclusion for the narrative.</p>	<p>Attempts to introduce a narrator and/or character(s). Attempts to organize an event sequence, but may have lapses. Attempts to use transition words to convey sequence and signal shifts from one time frame or setting to another. Provides a basic conclusion that follows from and reflects on the narrated experiences or events.</p>	<p>Logically introduces a narrator and/or character(s). Organizes an event sequence that unfolds naturally and logically. Effectively uses a variety of transition words to convey sequence and signal shifts from one time frame or setting to another. Provides a logical conclusion that follows from and reflects on the narrated experiences or events.</p>	<p>Skillfully introduces a narrator and/or character(s). Enriches the narrative by skillfully organizing an event sequence that unfolds naturally and logically. Consistently uses a variety of transition words to convey sequence and signal shifts from one time frame or setting to another, creating a strong cohesion throughout the narrative. Provides a compelling conclusion that enriches the narrated experiences or events.</p>
<p>Language/Conventions</p>	<p>Lacks a narrative style appropriate to task, purpose and audience. Uses simple or imprecise word choice. Makes significant errors in the conventions of Standard English grammar, usage, spelling, capitalization and punctuation which interfere with understanding the writing.</p>	<p>Attempts to establish a narrative style appropriate to task, purpose and audience. Occasionally uses precise words and sensory language in an attempt to capture the action and convey experiences and events. Makes frequent errors in using the conventions of Standard English grammar, usage, spelling, capitalization and punctuation which may interfere with understanding the writing.</p>	<p>Establishes and maintains a narrative style appropriate to task, purpose and audience. Effectively uses precise words and sensory language to capture the action and convey experiences and events. Effectively uses the conventions of Standard English grammar, usage, spelling, capitalization and punctuation with minor errors that do not interfere with understanding the writing.</p>	<p>Establishes and maintains an engaging narrative style which enhances the task, purpose and audience. Skillfully uses precise words and sensory language to capture the action and convey experiences and events, enriching the narrative. Skillfully uses the conventions of Standard English grammar, usage, spelling, capitalization and punctuation with few, minor errors that do not interfere with understanding the writing.</p>

Source: Created by the Kentucky Department of Education. Used with permission.

directions combined with rubrics that include explicit scoring criteria help limit the influence of subjectivity and unconscious bias (Brookhart, 2013; Panadero & Jonsson, 2020; Quinn, 2020a, 2020b), teachers must be constantly vigilant to ensure accuracy and fairness in their scoring.

Second, completing compositions and writing samples requires significant time on the part of students. They must gather background information, analyze its relevance, outline, and finally prepare their response. Advocates argue that these activities are useful learning experiences by themselves. Critics counter that these tasks take time away from other learning activities that may be equally valuable (Rahman, 2017). To resolve this problem, many teachers have students complete most of the planning and development work outside of class as homework. For this to be effective, however, teachers must ensure students have adequate resources at home, include explicit directions to relevant tasks, and then check the homework to offer students specific guidance on what improvements are needed. When teachers provide these types of support for homework, research indicates it can be highly effective (Cooper, 1989, 2007).

A third and probably the most serious drawback to using compositions and writing samples is the significant time they require for teachers to read, score, and offer feedback. Although developing rubrics with detailed scoring criteria makes this process much more efficient, it still involves considerable time. Imagine, for example, the average high school teacher who may see 120 students in class each day. If each of those students completes a composition, and if the teacher spends just *five minutes* reading each composition and offering students individualized feedback, the task will require *ten hours!*

Many mastery learning teachers resolve this problem by printing their rubric and scoring criteria for compositions or writing samples on a separate scoring guide. They then distribute two copies of the scoring guide to every student in the class and explain how to use it. Since these scoring criteria were presented and discussed as part of the learning unit, students should already be familiar with them. Next, students exchange the initial drafts of their papers as formative assessment A. Each student reads the work of a classmate and marks the scoring guide accordingly, being sure not to make any marks on the paper. When all students finish, they exchange papers again and conduct a similar review on a different classmate's paper. In this way, every student reads and evaluates the work of two different classmates. Finally, papers are returned to the author, with the two scoring guides attached. Students review the scoring guides for their paper, confer with their classmates if they note any discrepancies or need further clarification, and plan their revisions accordingly. They then revise their papers and submit the improved version to the teacher as formative assessment B.

This procedure offers significant benefits to both students and teachers. For students, reading the work of their classmates and using the scoring criteria can be a valuable learning experience. It gives them insights into how other

students approach writing and how the scoring criteria apply to their own work. In addition, the feedback they receive from their classmates is typically more detailed than that provided by the teacher, who must read not two but more than one hundred papers (Wu & Schunn, 2021). For teachers, they need to read each student's composition or writing only once, and then it's a revised version that is likely to be of fairly good quality. In Chapter 7, we discuss additional means such as screencasting (Morris & Chikwa, 2014; Séror, 2013) that mastery learning teachers use to make the feedback and corrective process easier and more efficient with these and other types of alternative assessments.

Laboratory Experiments

Teachers at every grade level use laboratory experiments as an alternative form of assessment. Although typically associated with science classes, laboratory experiments serve valuable instructional purposes in a wide variety of subject areas. Social studies teachers, for example, often engage students in experiments that involve the collection of survey data or polling information. These sorts of activities engage students in inquiry tasks that help develop skills such as observing, recording, and inferring. In addition, they allow students to construct their own understanding of phenomena and relationships.

Like other alternative assessment formats, laboratory experiments are best used to assess higher levels of cognitive skills. Teachers often combine laboratory experiments with traditional assessment formats within a single formative assessment experience to obtain better diagnostic information on student learning. In fact, the nature of laboratory experiments typically requires a multifaceted assessment. In most cases, they involve the demonstration of certain skills, lab notebooks in which students record their procedures and data, and formal reports on findings and interpretations prepared by students individually or in collaborative teams.

Creating good laboratory experiments that serve these valuable instructional purposes can be challenging, however. The process takes time and requires considerable scientific and technological know-how (Moeed, 2013; Shavelson & Baxter, 1992). Developing scoring criteria for laboratory experiments can be challenging as well, especially when teachers use a multifaceted approach. The procedures we described earlier concerning skill demonstrations, task performances and complex problems, and compositions can serve as useful guidelines in this process.

Projects and Reports

Projects and reports represent yet another common alternative assessment format used by teachers at all grade levels. Projects and reports require students to produce a product that demonstrates not only their understanding of important concepts but also their ability to organize and present that

understanding in meaningful ways. Projects are the primary assessment format in visual, commercial, and industrial arts and are used extensively by science and social studies teachers as well. They also provide the foundation for *project-based learning* (Kokotsaki et al., 2016; Lenz & Larmer, 2020). Reports are similarly used by teachers in every subject area and are especially common in integrated or thematic units of instruction.

In some cases, projects and reports become more comprehensive and may involve instruction over several units. For this reason, many mastery learning teachers use projects and reports as summative rather than as formative assessments. They divide the project or report into two or three parts with each part accomplished as part of a unique learning unit. A science teacher, for example, may divide the preparation of a research report into three parts: (1) a description of the problem and its importance, (2) methods and procedures used for the investigation, and (3) results and conclusions. As students complete each part, the teacher offers specific feedback to guide any needed corrective work, similar to what other teachers might do with compositions and writing samples. After correcting and refining each part, students assemble the parts in their project or report and submit them together as a summative assessment.

The key to success in using projects and reports as a culminating demonstration of what students have learned over a series of units rests in ensuring they provide direct evidence of students' mastery of specific learning goals. Sometimes teachers become enamored with the appeal of particular projects and lose sight of the purpose. Projects not directly aligned with the specified learning goals make poor assessments, regardless of how enticing or engaging they may be.

As for all alternative assessment formats, clearly articulated scoring criteria are vital to the successful use of projects and reports as formative or summative assessments. These criteria must be communicated to students as part of the instruction and presented with clear and concise examples based on students' work.

Portfolios

A *portfolio* presents a collection of a student's work. The term derives from artists' portfolios, which are collections of work put together by artists to show their style and range. Student portfolios have the same basic purpose: They are collections of students' products or performances that demonstrate their accomplishments or improvements over time. But portfolios are more than simple folders containing students' assignments. Rather, they are purposefully selected samples of students' work intended to show growth and development toward important learning goals (Lam, 2017; Russell & Airasian, 2011). Today, increasing numbers of teachers are using digital portfolios to display both visual and auditory content (Abrami et al., 2005; Cleveland, 2018; Renwick, 2017).

Because of the portfolios' broad scope, multiple components, and cumulative nature, most mastery learning teachers use them for the purposes of summative assessment rather than formative. Like projects and reports, portfolios represent culminating displays of students' accomplishments. Although each entry in the portfolio may have been developed through a process that included a series of formative assessments (see Clarke & Boud, 2018), the portfolio itself is more frequently used as a summative assessment.

Group Tasks or Activities

Most teachers today recognize the important opportunities and valuable benefits of cooperative learning (Gillies, 2016; Johnson & Johnson, 2002). In cooperative learning activities, students work together in small groups on collaborative projects or tasks to accomplish shared learning goals. These projects and tasks are designed to maximize the learning of *all* students in the group. Well-designed group tasks and activities require students to discuss the accuracy, quality, and quantity of their own and each other's work (Baloche & Brody, 2017; Johnson & Johnson, 1999). As a result, students not only master academic goals but develop important interpersonal social skills as well.

For group tasks and activities to be beneficial to students individually *and* to all other group members, they must include components for *individual accountability* and *positive interdependence*, sometimes referred to as *group responsibility* (Johnson & Johnson, 1995). That is, the performance of each student must be assessed individually, and the results given back to the student and to the group. In addition, students must recognize they cannot succeed in the task or activity unless the other members of their group also succeed.

Like other forms of alternative assessment, group tasks or activities require clear and unambiguous scoring criteria. For highly complex tasks or activities, separate criteria may be required for the group's work overall and for each student's individual contribution. This is true, for example, when teachers use the jigsaw method of cooperative learning. In jigsaw activities, a different portion of the learning task is assigned to each of five or six members on a team, and task completion requires contingent and mutual cooperation (Aronson et al., 1978; Karacop & Diken, 2017).

The particular criteria specified depend largely on the nature of the tasks or activities. The criteria described earlier under laboratory experiments, projects, and reports should be especially helpful in this effort. If the task or activity involves both academic and interpersonal social skill goals, separate criteria for each of these areas must be identified and clearly communicated to students. Furthermore, for highly complex tasks or activities that involve students working together for an extended period of time, mastery learning teachers typically divide the task into two or three subtasks. The assessment of each separate subtask then serves as a formative assessment, while evaluation of the overall task or activity is considered the summative assessment.

REFERENCES

- Abrami, P., Barrett, H., & Barrett, H. (2005). Directions for research and development on electronic portfolios. *Canadian Journal of Learning and Technology / La Revue Canadienne de l'apprentissage et de La Technologie*, 31(3). <https://doi.org/10.21432/T2RK5K>
- Aronson, E., Stephan, C., Sikes, J., Blaney, N., & Snapp, M. (1978). *The jigsaw classroom*. Sage.
- Arter, J., & McTighe, J. (2001). *Scoring rubrics in the classroom: Using performance criteria for assessing and improving student performance*. Corwin.
- Baloche, L., & Brody, C. M. (2017). Cooperative learning: Exploring challenges, crafting innovations. *Journal of Education for Teaching*, 43(3), 274–283. <https://doi.org/10.1080/02607476.2017.1319513>
- Brookhart, S. M. (2013). *How to create and use rubrics for formative assessment and grading*. Association for Supervision and Curriculum Development.
- Casner-Lotto, J., & Barrington, L. (2006). *Are they really ready to work? Employers' perspectives on the basic knowledge and applied skills of new entrants to the 21st-century U.S. workforce*. Partnership for 21st-Century Skills.
- Chappuis, J., & Stiggins, R. J. (2017). *An introduction to student involved assessment for learning* (7th ed.). Pearson.
- Clarke, J. L., & Boud, D. (2018). Refocusing portfolio assessment: Curating for feedback and portrayal. *Innovations in Education and Teaching International*, 55(4), 479–486. <https://doi.org/10.1080/14703297.2016.1250664>
- Clauser, B. E. (2000). Recurrent issues and recent advances in scoring performance assessments. *Applied Psychological Measurement*, 24(4), 310–324. <https://doi.org/10.1177/01466210022031778>
- Cleveland, R. E. (2018). Using digital portfolios: Reflection, assessment & employment. *TechTrends*, 62(3), 276–285. <https://doi.org/10.1007/s11528-018-0262-0>
- Cooper, H. (1989). Synthesis of research on homework. *Educational Leadership*, 47(3), 85–91.
- Cooper, H. (2007). *The battle over homework: An administrator's guide to setting sound and effective policy* (3rd ed.). Corwin.
- Gillies, R. M., & University of Queensland. (2016). Cooperative learning: Review of research and practice. *Australian Journal of Teacher Education*, 41(3), 39–54. <https://doi.org/10.14221/ajte.2016v41n3.3>
- Graham, S., Kiuahara, S. A., & MacKay, M. (2020). The effects of writing on learning in science, social studies, and mathematics: A meta-analysis. *Review of Educational Research*, 90(2), 179–226. <https://doi.org/10.3102/0034654320914744>
- Gronlund, N. E. (1993). *How to make achievement tests and assessments* (5th ed.). Allyn & Bacon.
- Harris, C. J., Krajcik, J. S., Pellegrino, J. W., & DeBarger, A. H. (2019). Designing knowledge-in-use assessments to promote deeper learning. *Educational Measurement*, 38(2), 53–67. <https://doi.org/10.1111/emip.12253>
- Johnson, D. W., & Johnson, R. T. (1995). Cooperative learning. In J. H. Block, S. T. Everson, & T. R. Guskey (Eds.), *School improvement programs* (pp. 25–56). Scholastic.
- Johnson, D. W., & Johnson, R. T. (1999). Making cooperative learning work. *Theory Into Practice*, 38(2), 67–73. <https://doi.org/10.1080/00405849909543834>
- Johnson, D. W., & Johnson, R. T. (2002). Learning together and alone: Overview and meta-analysis. *Asia Pacific Journal of Education*, 22(1), 95–105. <https://doi.org/10.1080/0218879020220110>
- Karacop, A., & Diken, E. H. (2017). The effects of jigsaw technique based on cooperative learning on prospective science teachers' science process skill. *Journal of Education and Practice*, 8(6), 86–97.
- Kokotsaki, D., Menzies, V., & Wiggins, A. (2016). Project-based learning: A review of the literature. *Improving Schools*, 19(3), 267–277. <https://doi.org/10.1177/1365480216659733>
- Lam, R. (2017). Taking stock of portfolio assessment scholarship: From research to practice. *Assessing Writing*, 31(1), 84–97. <https://doi.org/10.1016/j.asw.2016.08.003>

- Lenz, B., & Larmer, J. (2020). Project-based learning that makes a difference. *Educational Leadership*, 77(6), 66–70.
- Linn, R. L., Baker, E. L., & Dunbar, S. B. (1991). Complex, performance-based assessment: Expectations and validation criteria. *Educational Researcher*, 20(8), 15–21. <https://doi.org/10.3102/0013189X020008015>
- Moeed, A. (2013). Science investigation that best supports student learning: Teachers' understanding of science investigation. *International Journal of Environmental and Science Education*, 8(4), 537–559. <https://doi.org/10.12973/ijese.2013.218a>
- Morris, C., & Chikwa, G. (2014). Screencasts: How effective are they and how do students engage with them? *Active Learning in Higher Education*, 15(1), 25–37. <https://doi.org/10.1177/1469787413514654>
- Palm, T. (2008). Performance assessment and authentic assessment: A conceptual analysis of the literature. *Practical Assessment, Research, and Evaluation*, 13(4), 1–11.
- Panadero, E., & Jonsson, A. (2013). The use of scoring rubrics for formative assessment purposes revisited: A review. *Educational Research Review*, 9(1), 129–144.
- Quinn, D. M. (2020a). Experimental evidence on teachers' racial bias in student evaluation: The role of grading scales. *Educational Evaluation and Policy Analysis*, 42(3), 375–392. <https://doi.org/10.3102/0162373720932188>
- Quinn, D. M. (2020b). How to reduce racial bias in grading. *Education Next*, 20(4). <https://www.educationnext.org/how-to-reduce-racial-bias-in-grading-research/>
- Rahman, M. S. (2017). The advantages and disadvantages of using qualitative and quantitative approaches and methods in language “testing and assessment” research: A literature review. *Journal of Education and Learning*, 6(1), 102. <https://doi.org/10.5539/jel.v6n1p102>
- Renwick, M. (2017). *Digital portfolios in the classroom: Showcasing and assessing student work*. Association for Supervision and Curriculum Development.
- Rios, J. A., Ling, G., Pugh, R., Becker, D., & Bacall, A. (2020). Identifying critical 21st-Century skills for workplace success: A content analysis of job advertisements. *Educational Researcher*, 49(2), 80–89. <https://doi.org/10.3102/0013189X19890600>
- Russell, M. K., & Airasian, P. W. (2011). *Classroom assessment: Concepts and applications* (7th ed.). McGraw-Hill.
- Séror, J. (2013). Show me! Enhanced feedback through screencasting technology. *TESL Canada Journal*, 30(1), 104. <https://doi.org/10.18806/tesl.v30i1.1128>
- Shavelson, R. J., Baxter, G. P., & Pine, J. (1991). Performance assessment in science. *Applied Measurement in Education*, 4(4), 347–362. https://doi.org/10.1207/s15324818ame0404_7
- Shavelson, R. J., Baxter, G. P., & Pine, J. (1992). Research news and comment: performance assessments: political rhetoric and measurement reality. *Educational Researcher*, 21(2), 22–27.
- Sprouse, J. L., & Webb, J. E. (1994). *The Pygmalion effect and its influence on the grading and gender assignment of spelling and essay assessments* [Master's thesis]. University of Virginia.
- Stiggins, R. J. (1994). *Student-centered classroom assessment*. Merrill.
- Sweedler-Brown, C. O. (1992). The effect of training on the appearance bias of holistic essay graders. *Journal of Research and Development in Education*, 26(1), 24–29.
- Szetela, W., & Nicol, C. (1992). Evaluating problem solving in mathematics. *Educational Leadership*, 49(8), 42–45.
- Wu, Y., & Schunn, C. D. (2020). The effects of providing and receiving peer feedback on writing performance and learning of secondary school students. *American Educational Research Journal*, 58(3), 492–526. <https://doi.org/10.3102/0002831220945266>

Retrieved from the companion website for *Implementing Mastery Learning, Third Edition* by Thomas R. Guskey. Thousand Oaks, CA: Corwin, www.corwin.com. Copyright © 2023 by Corwin Press, Inc. All rights reserved. Reproduction authorized for educational use by educators, local school sites, and/or noncommercial or nonprofit entities that have purchased the book.