CHAPTER

Extended Written Response Assessment

We have all experienced extended written response exercises on tests. They may even have provided some of your best horror stories about your own assessment experiences. "Discuss Spain." "Analyze *King Lear*." "Explain the causes of the Civil War." Extended written response assessment requires students to construct a written answer, at least several sentences in length, in response to a question or task.

How can we write extended written response exercises so that all students have a fair chance of showing what they know? In this chapter, we will address issues of quality and of student involvement with extended written response assessment.[1] We will focus on the following topics:

- When to use the extended written response method.
- How to develop exercises and scoring mechanisms.
- How to involve students in planning, developing, and using extended written response assessment.

In this chapter, we focus on the shaded areas in Figures 6.1 and 6.2.

Figure 6.1 Keys to Quality Classroom Assessment

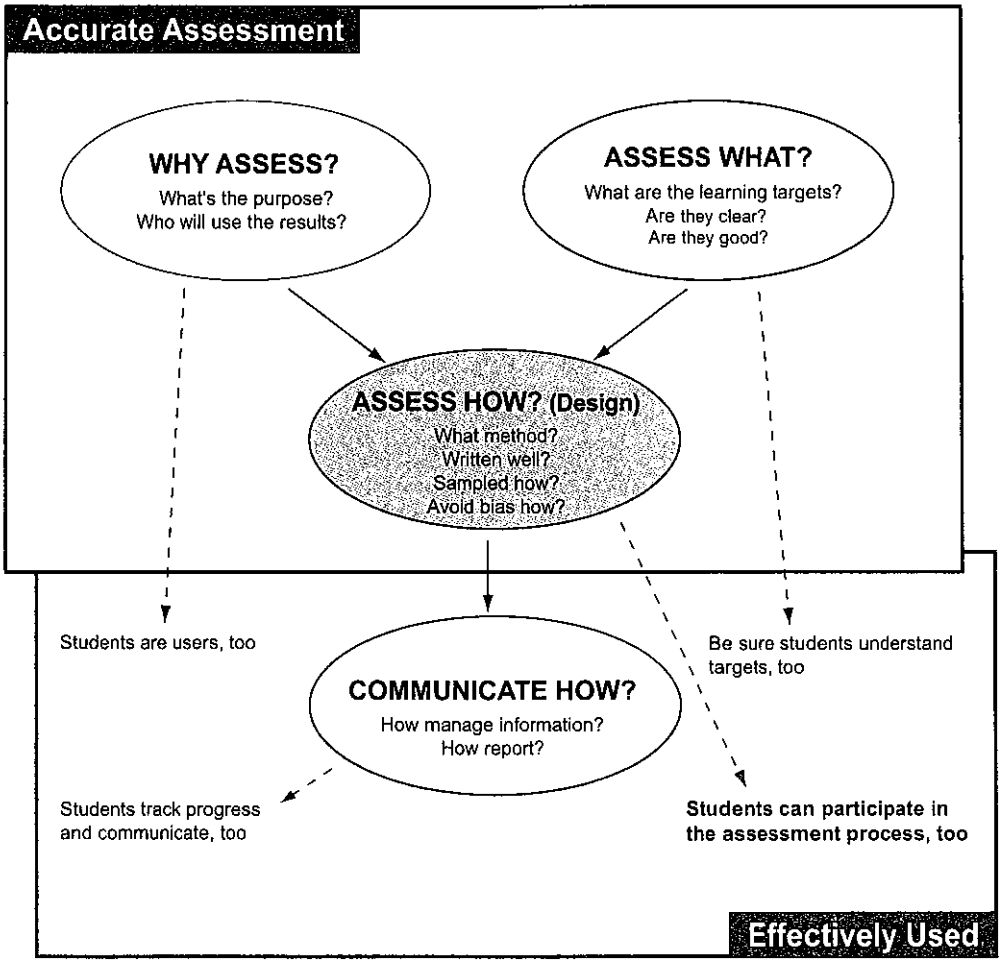


Figure 6.2 A Plan for Matching Assessment Methods with Achievement Targets

Target to Be Assessed	Assessment Method			
	Selected Response	Extended Written Response	Performance Assessment	Personal Communication
Knowledge Mastery				
Reasoning Proficiency				
Performance Skills				
Ability to Create Products				

Source: Adapted from *Student-Involved Assessment* for *Learning*, 4th ed. (p. 116), by R. J. Stiggins, 2005, Upper Saddle River, NJ: Merrill/Prentice Hall. Copyright © 2005 by Pearson Education, Inc. Adapted by permission of Pearson Education, Inc.

When to Use Extended Written Response Assessment

As we saw in Chapter 4, each individual classroom assessment represents a part of a bigger picture represented by a long-term assessment map that parallels the curriculum map. The kinds of learning targets to be assessed, combined with consideration of context factors, will determine whether some of those assessments will take the form of extended written response.

Matching Method to Target

As with all assessment methods, the first condition for using selected response is the type of learning target to be assessed, as described in Activity 4.2, "Target–Method Match," in Chapter 4. Knowledge and reasoning are ideal for extended written response exercises, as shown in Figure 6.2.

As described in Chapter 4, extended written response works well for assessing chunks of knowledge that interrelate, rather than individual pieces of knowledge assessed separately. (Selected response is more efficient for the latter.) For example, in science, we might want students to explain how atoms combine to form other substances, or to describe the Krebs cycle and explain why it is important. In social studies, we might want students to describe the factors that lead to placement of centers of habitation, and why each is important; for example, access to sources of transportation, water, food, and other natural resources.

Extended written response also works well for assessing reasoning. We can't open students' heads to look at their reasoning directly, but we can ask them to write down their thinking or rationale for a response. For example, in mathematics we might ask students to explain how they arrived at an answer. In science, we might ask students to explain their rationale for setting up an experiment. In addition, we can ask students directly to analyze, compare, make inferences about, and/or evaluate information. For example, in a unit on pollution, we might ask students to evaluate which solution to a problem is most likely to have the greatest benefit and to explain the reason for their choice.

TRY THIS

Activity 6.1 Learning Targets Best Assessed with Selected Response

Find three or four knowledge and reasoning learning targets in the curriculum materials you use that could be assessed by means of extended written response. Write them down and save them for later use.

Other Contextual Conditions

Several other conditions influence the selection of the extended written response method of assessment:

- Students need to be proficient in writing English. Extended written response may not work very well for primary students, English language learners, and students with other special needs.
- Extended written response can be time consuming to score well. If you can get the information you need (that is, reflect the target) through the less time-consuming selected response method, then do so.
- Extended written response exercises require consistency in scoring. Use this method only when you know that the scoring guides are of good quality and that scorers will apply them consistently.
- Extended written response is good in assessment *for* learning situations, especially when students are analyzing the quality of reasoning in their own work.

Developing Extended Written Response Assessments

Designing and developing extended written response assessments entail the same five stages, introduced in Chapter 4, that we used with selected response assessment:

1. Plan the assessment.
2. Develop the assessment.
3. Critique the assessment.
4. Administer the assessment.
5. Revise the assessment.

Stage 1: Plan the Assessment

At this first stage, as you remember, we answer four questions: Assess why? Assess what? Assess how? How important? Figure 6.3 presents the range of answers to these questions.

Figure 6.3 The Planning Stage for Extended Response Assessment

Assess why?

Assessment *for* Learning

- Diagnostic/planning information for teacher
- Diagnostic/planning information for student
- Diagnostic/planning information for others
- Student self-evaluation

Assessment *of* Learning

- Report card grade

Assess what?

Knowledge, reasoning, and/or writing learning targets

Assess how?

Extended response exercises and scoring devices

How important?

Sample of items to represent the breadth of the learning targets in question and their relative importance in the learning plan

Assess Why?

As you recall, we build assessments to meet specific user information needs. Thus, in every case, your number one decision is, Will this be an assessment *for* learning or an assessment *of* learning? Who will be using the assessment results and how will they be using them? If it is an assessment *for* learning, is it purely for your use in instructional decision making or is it to serve your students, too? Obviously, this will influence how the assessment relates to your teaching and how you construct, administer, and share the information from the assessment.

Assess What?

For extended written response assessments, we outline the content to be mastered, including the relationships among elements that we expect our students to know and understand. If you wish students to demonstrate mastery of patterns of reasoning, what patterns? Your list of content knowledge, understanding, and reasoning provides the ingredients for exercise development.

Assess How?

Verify one final time that these targets are appropriate for transforming into the extended written response format.

How Important?

Again, it's important to establish priorities at the outset. Which of the learning targets or topics are most important, next most important, and so on? This will serve as the basis for the distribution of points or ratings in the overall assessment plan. The prioritization should parallel the amount of time and emphasis given the various targets or topics in teaching.

As described in Chapter 4, we can create a plan by using either a table or a list of learning targets.

Plans for extended written response tests are similar in their basic framework to those used for selected response assessments, and are different in other ways. Table 6.1 (a repeat of Chapter 1, Table 1.1) is an example of a table of specifications for an extended written response test covering a unit on pollution. Mr. Heim listed the categories of information students are to know on one axis and the patterns of reasoning they are to master on the other. Row and column totals, and therefore entries in the cells of the table, represent the relative emphasis assigned to each.

Unlike tables of specifications for selected response assessments, the cells in the extended written response test plan in Table 6.1 contain the number of points on the test assigned to that content-reasoning combination, not the number of individual test items. Given 50 points for the entire exam, this plan emphasizes how to reduce pollution, requiring that students rely on that understanding to compare and evaluate.

Table 6.1 Mr. Heim's Sample Test Blueprint

		PATTERN OF REASONING		
	Know	**Compare**	**Evaluate**	**Total**
Concentrations	10	0	0	10
Effects of Pollutants	7	8	0	15
How to Reduce Pollution	6	10	9	25
Total	23	18	9	50

TRY THIS

Activity 6.2 Create an Extended Written Response Assessment

Select one of the learning targets you identified in Activity 6.1. Draft an extended written response exercise and a scoring scheme. Put them aside for later use.

Stage 2: Develop the Assessment

One of the advantages often listed for extended written response tests relative to other test formats is that exercises are much easier and less time consuming to develop. Keep in mind, however, that "easier to develop" does not mean they require little thought.

To succeed with this assessment format, we must first write exercises that describe a single, complete, and novel task. Second, we must devise clearly articulated evaluation criteria. If we are not careful at this stage, students who know the material may not perform well, and students who have not mastered the material may be able to look as though they have. Poorly framed extended written response exercises can be a nightmare for students to answer and for teachers to score.

Devising Exercises

Sound extended written response exercises do three things: (1) set a clear and specific context; (2) specify the kind of reasoning to be brought to bear; and (3) point the way to an appropriate response without giving away the store.

Setting the Context

Set the context in the exercise by specifying the knowledge to be brought to bear. For example:

> *During the term, we have discussed both the evolution of Spanish literature and the changing political climate in Spain during the twentieth century.*

Specifying the Reasoning

Specify the kind(s) of reasoning or problem solving, if any, students are to carry out. For example:

> *During the term, we have discussed both the evolution of Spanish literature and the changing political climate in Spain during the twentieth century. Analyze these two dimensions of life in Spain, citing instances where literature and politics may have influenced each other. Describe those influences in specific terms.*

Pointing the Way

Point the direction to an appropriate response by reminding students of the criteria that will be applied in evaluating responses. For example:

> *During the term, we have discussed both the evolution of Spanish literature and the changing political climate in Spain during the twentieth century. Analyze these two dimensions of life in Spain, citing instances where you think literature and politics may have influenced each other. Describe the influences in specific terms. In planning your response, think about what we learned about prominent novelists, political satirists, and prominent political figures of Spain. (5 points per instance, total = 15 points.)*

Let's analyze an extended written response exercise that might arise from a test covering the content of this book on classroom assessment as applied to foreign-language speaking proficiency:

Assume you are a French teacher with many years of teaching experience. You place great value on the development of speaking proficiency as an outcome of your instruction. Therefore, you rely heavily on assessments where you listen to and evaluate performance. But a problem has arisen. Parents of students who attained very high scores on your written tests are complaining that their children are receiving lower grades on their report cards. The principal wants to be sure your judgments of student proficiency are sound and so has asked you to explain and defend your procedures. Describe at least three specific quality standards that your oral proficiency exams would need to meet for you to be confident that your exams truly reflect what students can do; provide the rationale for each. (2 points for each procedure and rationale, total = 6 points.)

Here's the challenge presented to students in a nutshell:

Demonstrate understanding of:	Performance assessment methodology
By using it to figure out:	Proper applications of the method in a specific context
Adhering to these standards:	Include three appropriate procedures and defend them

Offering Choices

We recommend that you don't offer choices. The summative assessment question should always be, "Can you hit the agreed-on target?" It should never be, "Which (or which part of the) target are you most confident that you can hit?" When students select their own sample of performance, it can be a biased one.

Interpretive Exercises

With this format, as with selected response, you don't have to assume that students always have mastered the content around which an exercise is built. If you wish to assess student mastery of specific patterns of reasoning only, you can provide them with a table, chart, or map of background information about a given topic and then ask them to write a response demonstrating that they can figure out relationships between or among things presented. For example, you might present a chart of data, a map, or even a passage of details and ask them to ferret out and depict certain relationships, draw comparisons, conduct analyses,

TRY THIS

Activity 6.3 Revise Your Exercise

Please return now to the extended written response exercise you developed in Activity 6.2. Did you specify the knowledge and kind(s) of reasoning students are to use? Did you remind them of standards by which their work will be judged? Revise your exercise as needed to meet these guidelines.

or create and fill categories. Scoring criteria for essays, then, would reflect the active ingredients of sound reasoning.

Developing Extended Written Response Scoring Procedures

When a student writes a response, we can evaluate whether the student's work demonstrates three different qualities:

- Accurate knowledge and understanding
- Sound reasoning
- Effective written communication

The first two focus on matters of content, while the last treats the response as a product. As we mentioned earlier, in this chapter we focus on how to evaluate the content of the response: accuracy of knowledge and quality of reasoning.

(When the characteristics of effective written communication are evaluated, we would classify the learning target as a product and evaluate it with a performance assessment. A rubric for evaluating writing quality, "6 + 1 Traits," is provided on the CD in the file, "Rubric Sampler.")

Not surprisingly, a key to successful use of extended written response assessment is the clear articulation of appropriate evaluation criteria by which to judge the quality of student responses. We recommend not using "floating standards," in which the evaluator waits to see what responses come in before deciding how to score. Floating standards destroy the validity and reliability of the assessment. Teachers and students alike need to be clear in advance regarding which aspects of the response are important.

Exercise-Specific Scoring

This is the typical procedure. We award points when specific information appears in students' answers. The French teacher example that appeared in the previous section calls for this kind of scoring. Here is the scoring guide:

> *Give two points if the student's response lists any of these six procedures and defends each as a key to conducting sound performance assessments:*
>
> - Specify clear performance criteria
> - Sample performance over several exercises
> - Apply systematic rating procedures
> - Maintain complete and accurate records
> - Use published performance assessments to verify results of classroom assessments
> - Use multiple observers to corroborate
>
> *Also award two points if the response lists any of the following and defends them as attributes of sound assessments:*
>
> - Specifies a clear instructional objective
> - Relies on a proper assessment method
> - Samples performance well
> - Controls for sources of rater bias
>
> *All other responses receive no points.*

Generic Rubrics

Although we often use exercise-specific point-based scoring guides for assessing student understanding of content, they are not our only option. We can use a generic rubric or scoring guide that defines what content understanding looks like, in general, for any body of knowledge. For example, a three-point generic rating scale for "understanding the content" might define three levels of mastery of the required material. Here's an example (a more detailed example of a general rubric for evaluating content understanding, "Essay Scoring Criteria," is provided on the CD in the file, "Rubric Sampler"):

> 3 The response is clear, focused, and accurate. Relevant points are made with good support. Good connections are drawn and important insights are evident. Vocabulary is used correctly.

2 The response is clear and somewhat focused, but not compelling. Support of points made is limited. Connections are fuzzy, leading to few important insights. Sometimes vocabulary is used correctly, sometimes not.

1 The response misses the point, contains inaccurate information, or otherwise demonstrates lack of mastery of the material. Points are unclear, support is missing, and/or no insights are included. Vocabulary is often used incorrectly.

Notice that such a scoring guide can be used to assign points to student understanding of any body of knowledge. This generic way of evaluating content understanding is especially useful when students are demonstrating their reasoning abilities using different bodies of knowledge. This frequently happens in performance assessments, portfolios, or individual research reports. For the type of understanding we typically assess through extended written response, however, we usually stick with exercise-specific, point-based scoring.

The situation is different if we're also scoring how well students reason with the knowledge given in the exercise. In this case, a generic rubric for the type of reasoning to be demonstrated is required. For example, if the extended written response exercise calls for students to make a generalization based on content, you might use an exercise-specific rubric to assess content understanding, and then a generic rubric to analyze the quality of the generalization, as shown in Figure 6.4.

We recommend generic rubrics for assessing the quality of various patterns of reasoning because they can be shared with students in advance without giving away the answer, used again and again, and overlaid on any body of content. Used this way, they help students come to learn the features of solid reasoning.

Stage 3: Critique the Assessment

An excellent way to check the quality of your exercises is to try to write or outline a high-quality response yourself. If you can, you probably have a properly focused exercise. If you cannot, it needs work.

Remember that there can be sources of bias specific to extended written response assessment. These arise, for example, if students are not yet proficient writers or are English-language learners. Figure 6.5 summarizes the factors to think about when devising

extended written response exercises and scoring procedures. Answering these questions assists in constructing effective, high-quality exercises—those that avoid bias and distortion.

Figure 6.4 Rubric for Generalization

Strong

- Statement is true for evidence presented and extends application logically to a broader array of instances.

Part-way There

- Statement is true for evidence presented, but application includes too broad an array of instances to be supported by evidence.
- Statement is true for evidence presented, but application to other instances is based on too little evidence.

Beginning

- Statement is true for evidence presented, but no extension beyond the evidence is attempted.
- Statement is true for evidence presented, but application to other instances is not related to evidence.
- Statement is not true for evidence presented.

TRY THIS

Activity 6.4 Check Your Scoring Criteria

Return to the scoring plan you developed for your exercise in Activity 6.2. Are you clear about what features of the response you will score—content, reasoning, and/or the quality of the writing? Revise your exercise as needed.

Figure 6.5 Quality Guidelines for Extended Written Response Assessments

Quality of the Exercises

- Is extended written response the best assessment method for this learning target?
- Do exercises call for brief, focused responses?
- Is the target knowledge clear?
- Is the reasoning to be done, if any, clear?
- Is the exercise itself written at the lowest possible reading level—will all students understand what they are to do?
- Will students' level of writing proficiency in English be adequate to show you what they know and can do?
- Is there anything in the question that might put a group of students at a disadvantage regardless of their knowledge or reasoning level?
- Are there enough exercises to provide a defendable estimate of student learning on intended targets?

Quality of the Scoring Guide(s)

- Would experts in the field agree with the definition of a quality response?
- For the knowledge aspect of the response, are the criteria clear—would the elements in the scoring plan be obvious to good students without giving away the answer?
- For the reasoning portion of the response (if any), is there a generic rubric that captures the essence of high quality thinking?
- Do the criteria match the exercise?

Scoring Considerations

- Will the number of students to be evaluated be such that the rater(s) can adequately assess each response?
- Is there an adequate amount of person time available to read and evaluate responses?
- Have all scorers been adequately trained to score the essays consistently—will there by a high level of rater agreement?

Source: Adapted from *Student-Involved Assessment for Learning*, 4th ed. (pp. 130, 135, 136), by R. J. Stiggins, 2005, Upper Saddle River, NJ: Merrill/Prentice Hall.

Stages 4 and 5: Administer the Assessment, Watch for Problems, and Revise as Needed

After you conduct, score and interpret the assessment, if it has flaws you will see them very clearly and can then correct them before future use. You also will see if your instruction has fallen short on particular standards, which allows you to fix that both with your current set of students and in future instruction.

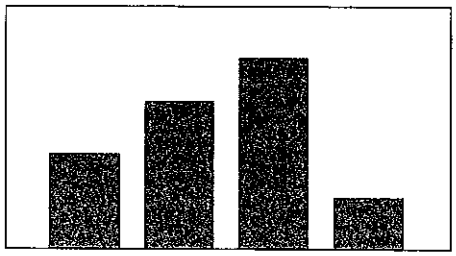
DEEPEN UNDERSTANDING

Activity 6.5 Analyze Extended Written Response Assessments for Quality

If you are working with a team, make sure everyone understands the quality guidelines in Figure 6.5 by using those guidelines individually to critique the sample essay questions in Figure 6.6 and then comparing critiques.

Figure 6.6 Sample Extended Written Response Exercises

Essay 1: Label the Graph.* This question is intended for grades 3–12. It is one of six exercises using different content to assess problem solving in mathematics. Results are used to track individual student progress toward mastery of state content standards. The scoring criteria have four traits, each scored separately by trained raters—conceptual understanding, mathematical procedures, strategic reasoning, and communication in mathematics. Students may or may not see the criteria depending on the teacher.



1. What might this be the graph of? Put titles and numbers on the graph to show what you mean.
2. Write down everything you know from your graph.

Essay 2: Assembly Line. Intended for grade 6. No scoring mechanism, uses, or targets are described.

"Describe the effect of the development of the assembly line on American society."

*Used in several assessments, for example the Oregon state mathematics assessment in 1996.

Figure 6.6 (Continued)

Essay 3: Day and Night. The authors provide examples of several different types of assessment tasks in science—short response to assess knowledge, extended response to assess conceptual understanding, and performance assessments to assess "science knowledge in context," integration of knowledge, practical skills, and reporting skills. The authors use all the sample questions as examples to illustrate methods and when to use them; they specifically state "the assessment tasks provided. . . are not exhaustive, but exemplary. . . ." The sample questions are not meant to be a complete test of any targets.

The following task is intended for grade 2 to assess science understanding.

"Everyone knows about day and night. Write what you think makes day and night. (Four primary lines are given for the response.) Draw a picture to show what you think. (5"x 5" box given for response.)"

The scoring criteria are shown in the following table. Students don't see the criteria.

Score	Label	Description	Examples
2	Scientific Conception	The response indicates that the Earth turns so that the same face is not always facing the Sun.	"The Earth turns every 24 hours and for 12 hours we are facing the Sun."
1	Opposite Sides	The response indicates that the Moon and Sun are on different sides of the Earth and the Earth rotates facing one and then the other. There is no implication that the Sun moves.	"In the day we face the Sun and in the night we turn to face the Moon."
0	Sun Moves	The response indicates that the Sun moves to cause night and day (possibly across the sky).	"The Sun moves and makes way for the Moon."
Additional Notes: Some responses may have mixed the elements of the last two categories. If so, give the score for the lower response. An indication that the Sun moves should take precedence in determining the lower category.			

Figure 6.6 (Continued)

Essay 4: Emerson Quiz. This quiz is intended for grades 10–12 to assess mastery of content knowledge (knowledge of Emerson) and reasoning in literature. Results will be used as 10 percent of the final grade in a literature class. Two of the ten essay questions are provided below. Students get 1 point for their answer and 1 point for their rationale.

"Read each of the statements below and put a check if Emerson would most likely complete the activity or put an X if he would disagree or not do the listed activity. For each answer, write your rationale. Include a statement from Emerson's work to support your check or X. Be sure to quote the statement directly and give the page number in parentheses. Use the introduction to Emerson, Nature, and 'Self-Reliance.'

2. ____ look to the past for guidance.
5. ____ join a popular civic organization."

Source: Adapted from *Practice with Student-Involved Classroom Assessment* (pp. 142–143), by J. A. Arter & K. U. Busick, 2001, Portland, OR: Assessment Training Institute. Copyright © 2006, 2001 Educational Testing Service.Exercise "Day and Night" reprinted from *Exemplary Assessment Materials—Science* (p. 15), by Australian Council for Educational Research Ltd., 1996, Victoria, NSW, Australia. Exercise "Emerson Quiz" reprinted from Thomas Mavor, 1999, Brother Martin High School, New Orleans, LA. Reprinted and adapted by permission.

Extended Written Response Assessment *for* Learning

Remember, student motivation and achievement both improve when we use the assessment process to help students answer the following three questions: "Where am I going?"; "Where am I now?"; and "How can I close the gap?" The same strategies we use with selected response tests, as summarized in Figure 5.8, also work with extended written response assessments.

For example, we can engage students in devising practice exercises like those that will appear on a future examination. This will help them learn to center on important content and will require that they become sufficiently comfortable with our valued patterns of reasoning that they can build them into practice exercises. If they write practice exercises, trade with classmates, and write practice responses, both we and they gain access to useful information on what parts of the standards they are and are not mastering.

For a simpler application of this idea, provide students with practice exercises and see if they can place them in the proper cells of the test plan. Then have them defend their placement.

Additionally, we can make use of some scoring guides as instructional tools. Exercise-specific scoring guides cannot function as teaching tools; students cannot use them to

practice scoring because these guides outline the specifics of acceptable responses to one exercise. However, you can involve students with developing sample scoring guides for those practice exercises they have created. Or, you can provide sample exercises and have students practice developing scoring guides for them. Further, they can practice scoring each other's responses to those exercises. By repeating this process as students proceed through a unit of study, you can provide them with opportunities to watch themselves improve. Chapter 7 contains more detailed examples of how generic rubrics can be used as instructional tools.

Summary

Extended written response assessments are excellent for assessing extended bodies of knowledge and reasoning learning targets. We followed the creation of these assessments through five stages, with an in-depth focus on the development stage. Exercises need to specify what knowledge and patterns of reasoning, if any, students are to use in crafting their response. They also need to indicate what features of performance will count, by pointing the way to the correct answer without giving away the store. Exercises must avoid other potential sources of bias and distortion such as unclearly written instructions, instructions at too high a reading level, and features that might disadvantage any group. Scoring procedures and guides must be developed along with the exercises. We explored two options: the exercise-specific scoring guide and the generic rubric. The first is most typically used to call out content knowledge that must be present in a correct response, while the second is useful for evaluating patterns of reasoning. We shared a generic rubric for content knowledge, as well.

We concluded with some suggestions for strategies that use extended written response items as assessment *for* learning, where students share the exercise development and scoring responsibility. These strategies connect assessment to teaching and learning in ways that can maximize both students' motivation to learn and their actual achievement.

Tracking Your Learning—Possible Portfolio Entries

Any of the activities included in Chapter 6 can be used as portfolio entries. Remember, the learning targets for this book are outlined in Figure 6.1, listed in Table 1.2, and described in detail in Chapter 1. The portfolio entry cover sheet provided on the CD in the file, "Portfolio Entry Cover Sheet," will prompt you to think about how each item you select reflects your learning with respect to one or more of these learning targets.

Either of the following activities would also make good portfolio entries. Each could also be used as a learning team activity.

TRY THIS

Activity 6.6 Analyze Your Own Extended Written Response Assessments for Quality

Find an extended written response assessment that came with a text or with the teaching materials you use. Evaluate the test, following the recommendations given in this chapter for critiquing an assessment. Write a short commentary detailing its strengths and weaknesses. If you plan to use the test, revise it as needed.

TRY THIS

Activity 6.7 Develop an Extended Written Response Assessment

Select an upcoming context (a unit, an area of study, a collection of related learning targets) for which an extended written response assessment would be the appropriate assessment. Create that assessment, following the guidelines described in this chapter.

Notes

1. Portions of this chapter have been reprinted and adapted from Chapter 5, pp. 115–139, of R. J. Stiggins, *Student-Involved Assessment* for *Learning*, 4th ed., 2005, Upper Saddle River, NJ: Merrill/Prentice Hall. Copyright © 2005 by Pearson Education, Inc. Reprinted and adapted by permission of Pearson Education, Inc.