

CHAPTER

7

Performance Assessment

[P] erformance assessment, once described as the most authentic of assessment methods, now is seen just as it should be—one valuable way to collect information about student achievement and to involve students in learning. As with any other assessment method, it can be constructed and used well or poorly. This chapter focuses on the following:

- What performance assessment is.
- Which learning targets can be assessed well through performance assessment.
- How to develop the two parts to a performance assessment—tasks and criteria.
- How to determine the quality of tasks and criteria.
- Strategies for using performance criteria as assessment *for* learning.

This chapter demonstrates how one assessment method, performance assessment, fits into each of the keys in Figure 7.1, and delves into the highlighted portion of the target-method match shown in Figure 7.2.

Figure 7.1 Keys to Quality Classroom Assessment

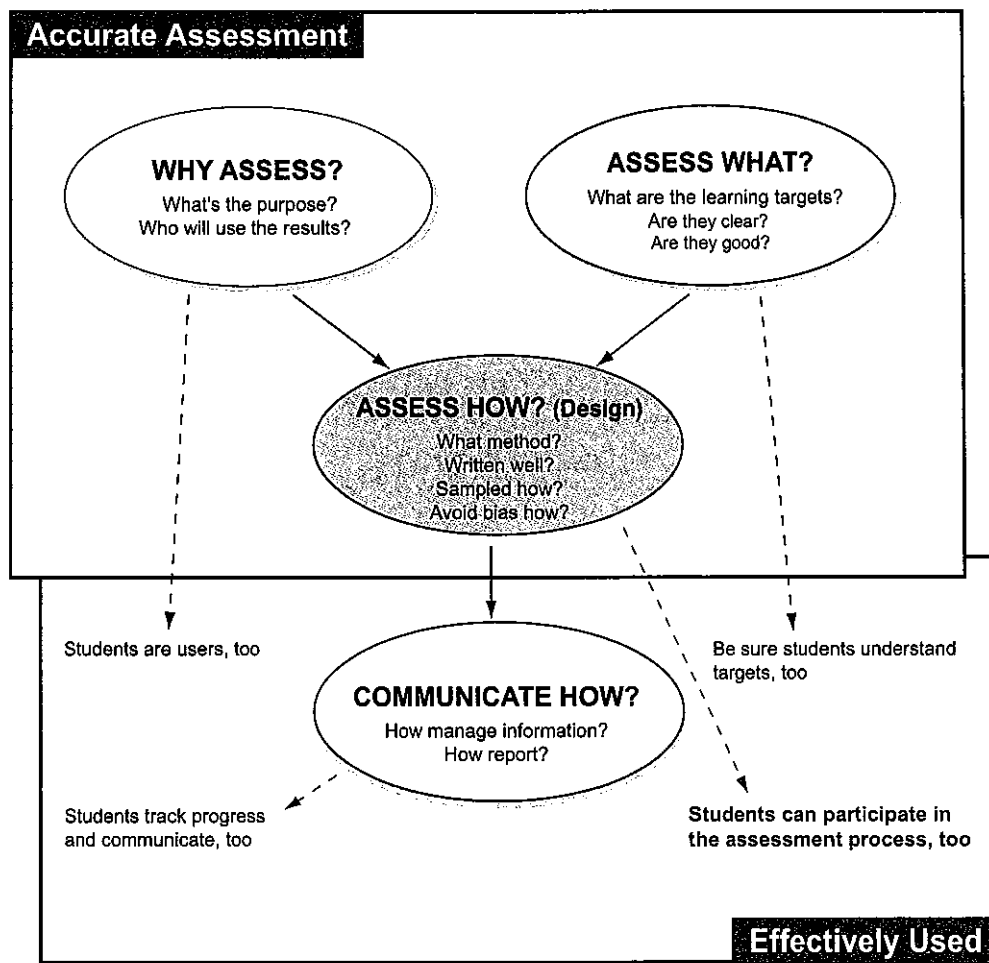


Figure 7.2 A Plan for Matching Assessment Methods with Achievement Targets

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. 141), 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.

Assessment Based on Observation and Judgment

Definition of Performance Assessment

Performance assessment is assessment based on observation and judgment. Students engage in an activity that requires them to apply a performance skill or create a product and we judge its quality. Examples appear in Table 7.1.

Humans have been using performance assessment for at least 3,000 years. To obtain a government position in ancient China, applicants had to demonstrate their skill at writing poetry and shooting with a bow and arrow. Although performance assessment is not a recent innovation, the following aspects of it are more recent developments.

Table 7.1 Examples of Performance Tasks and Performance Criteria

Example	Task(s)	Performance Criteria
Selection onto the wrestling team	Timed run, sit-ups, pull-ups, exhibition matches	Physical fitness; knowledge of the moves; motivation
Old-time fiddle competition	Play three songs—hoe-down, waltz, and tune of choice—total not to exceed three minutes	Quality of intonation, phrasing, and creativity; precision of playing; avoidance of certain bowing patterns
Judging art	Paint a picture, make a sculpture	Quality of theme, composition, use of color
Reading comprehension	Read the story <i>Stone Fox</i> and respond to prompts such as, "Record your thoughts, feelings, and questions about what you are reading"; "Select a line or lines from the story that interest you or make you think and tell why you chose these lines."	Score the quality of the responses on a 6-point scale where 6 is an "exemplary" reading performance characterized by insight, constructing and reflecting on meaning in a text, developing connections within and among texts, taking risks, and challenging the text.
Science process skills	Students circulate to six different stations to do hands-on science activities such as analyzing seeds to judge how they might be dispersed. The purpose is to see how well students can make observations, draw inferences from those observations, and use this information to predict what will happen with other plants.	Responses are scored on a scale of 1-4 in two areas: <i>problem solving/inquiry</i> and <i>communication</i> . A "4" performance on problem solving means that the student analyzed and readily understood the task, developed an efficient and workable strategy, effectively implemented the strategy, and appropriately applied critical knowledge.
Reasoning and persuasive writing	"Imagine that it is 1858 and you make a special trip to hear Abraham Lincoln and Stephen Douglas debating during their campaigns for the Senate. Write an essay in which you explain the most important ideas and issues."*	Students receive scores in content, use of relevant previous knowledge, use of general social studies concepts, quality of the argumentation, appropriate reference to reading materials supplied, and number of historical misconceptions.

* Performance assessment task developed by the National Center for Research on Evaluation, Standards, and Student Testing, UCLA, 300 Charles E. Young Dr. N, mailbox 951522, Los Angeles, CA, phone 310-206-1532; website <http://www.cse.ucla.edu>.

- *Systematization.* Performance assessment is essentially subjective. Our challenge is to make this subjective form of assessment as objective (free of bias and distortion) as possible by making it more systematic. A lot of the work of making performance assessment systematic is in the area of creating high-caliber mechanisms for rating the quality of performance and teaching human raters to use the criteria to judge performances consistently.
- *Large-scale uses.* The work on consistency has enabled performance assessment to be used more formally outside the classroom as part of standardized, high-stakes tests.
- *Assessment for learning.* The work on consistency, oddly enough, has also led to performance assessment's preeminence in involving students in assessment by directing them to the features of work that constitute a quality performance and enabling them to judge levels of quality for themselves.

The Two Parts of a Performance Assessment

Tasks

Tasks are the activities or exercises in which students engage while the teacher observes and judges quality. For example:

- The tasks when being assessed for a driver's license consist of controlling the car under various conditions—driving on the freeway, driving on one- and two-way streets, stopping at stop signs, negotiating turn lanes, backing around a corner, parallel parking, and so on.
- The tasks for Olympic diving consist of a certain number of prescribed dives and a certain number of elective dives.
- The task for judging oral presentation skills is to give an oral presentation.
- The task for ability to work in a group is to have students work in a group.
- The task for judging the ability to cook is to have students cook a variety of dishes.
- The task for judging the ability to carry out experiments in science is to have students conduct a variety of experiments.

A *task* is any activity that we use as a context to observe a skill or a product—a naturally occurring event or a separate event, at the end of instruction or during instruction. The only requirement is that the task elicits the desired skill or product so that it is capable of being judged.

Performance Criteria

What words come to mind when you think of “criteria?” Standards for comparison? Benchmarks? What counts? What you’re looking for in work? These are all terms associated with the word. *We define performance criteria as the basis for judging the quality of the performance on the task.* Here are some examples:

- The criteria for the driver’s test are the extent to which candidates obey the law, do not cause congestion, and do not cause a danger.
- The criteria for Olympic diving include difficulty of the dive, form, and amount of splash when the diver enters the water.
- The criteria for an oral presentation include the content of what is presented, how the presentation is organized, the appropriateness of the language used for the audience, and the delivery.
- The criteria for group work could include respect for others, degree of participation, and quality of the ideas contributed.
- The criteria for cooking a variety of dishes include taste, texture, scent, and presentation.
- The criteria for scientific experimentation might involve such things as reasoning through the variables that must be controlled, setting up the equipment in the needed configuration, collecting the information needed in an accurate manner, accurately recording data, and drawing a sound conclusion from the information collected.

We will use the words *performance criteria*, *rubrics*, *assessment lists*, and *scoring guides* interchangeably in this chapter. Just be aware that not everyone who writes about performance assessment uses these words to mean the same thing. For our purposes here and in the classroom, criteria can be cast in the form of points assigned for specific features of a response, a checklist of important features, or a rating scale. What’s impor-

tant is that the criteria, whatever we call them, present a clear and accurate picture of what constitutes quality.

When to Use Performance Assessment

In Chapter 4 we made the case for reserving performance assessment for those learning targets that really need it: some forms of reasoning, performance skills, and products. Table 7.2 lists concrete examples of such targets for various grade levels from the content standards of several states. Clearly, the most important indicator for use of performance assessment is the nature of the learning target, but other factors may argue for its use as well:

- *The age of students.* About the only ways to determine what younger primary students know and can do are either to ask them about it orally (personal communication) or to watch them do something (performance assessment).
- *Reading and writing proficiency.* Other assessment methods might not provide accurate information in certain contexts, such as with English language learners, or students who don't read or write sufficiently well to answer test questions.

Developing Performance Assessments

The stages in developing performance assessments are those described in Chapter 4 for all assessment methods:

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

Table 7.2 Examples of Learning Targets Assessable Using Performance Assessment

Reading	<p>Reasoning: Evaluate the validity of what is read.</p> <p>Skill: Read at a rate of 110 words per minute by the end of grade 2.</p>
Writing	<p>Product: Vary form, detail, and structure of writing in accordance with intended audience and purpose.</p> <p>Product: Use language that is precise, engaging, and well suited to the topic and audience.</p>
Communication	<p>Skill: Communicate using different forms of oral presentation.</p> <p>Skill: Use effective listening and speaking behaviors.</p>
Mathematics	<p>Reasoning: Recognize when an approach is unproductive and try a new approach.</p> <p>Skill: Accurately measure temperature, distance, weight, and height.</p> <p>Reasoning/Skill: Support a conclusion or prediction orally and in writing, using information in a table or graph.</p> <p>Product: Create three-dimensional objects.</p>
Physical Science	<p>Skill: Choose and use laboratory equipment properly to design and carry out an experiment.</p>
History	<p>Product: Organize and record information.</p>
Geography	<p>Product: Use data and a variety of symbols and colors to create thematic maps and graphs (e.g., patterns of population, rainfall, or economic features).</p>
Social Studies	<p>Reasoning and Skills: Give examples of and exhibit the behavior of good classroom citizens, including respect, kindness, self-control, cooperation, sharing common resources, and good manners.</p> <p>Skill: Use maps and globes to locate places referenced in stories and real-life situations.</p>
The Arts	<p>Reasoning: Compare and contrast artwork in terms of elements of design.</p> <p>Product: Organize art elements to develop a composition and to change the impact of a composition.</p>
Health/Fitness	<p>Skill: Wrestle.</p> <p>Product: Create and implement a health and fitness plan.</p>
Shop	<p>Reasoning: Diagnose car engine problems.</p> <p>Product: Make a functional object out of wood or metal.</p>

TRY THIS**Activity 7.1 Learning Targets Best Assessed with Performance Assessment**

List several specific learning targets from your current assignment that should be assessed using performance assessment. (Don't list everything that could be assessed this way, just those that are of the highest priority.) Save this list for future use.

Stage 1: Plan the Assessment

Plan the assessment by answering four questions: Assess why? Assess what? Assess how? How important?

Assess Why?

Again, we begin with consideration of how the results of the assessment are to be used and by whom. What decisions will the information drive? How will the assessment be used to meet students' information needs? Specific design choices we make are guided by answers to these questions.

For example, rubrics for use with students will look different than those used on a large-scale, standardized assessment. Assessments developed to provide descriptive feedback may look different from those designed to provide an overall picture of student achievement status.

Assess What?

This question requires that we specify the learning target or targets the assessment is to measure. Although this step may seem simple, it becomes important to have done it when we get to selecting or creating criteria, so don't skip it.

Assess How?

At this step in planning, we double check that performance assessment is the best choice for our purpose and our targets.

How Important?

Have you ever had the feeling that the single state writing sample didn't accurately represent a student's true ability to write? Perhaps the topic was familiar to the student and she performed better than usual, or the task was worded in such a manner that the student might not have understood it and done worse than usual. Or, maybe you noticed that a single event, such as having a student read out loud, didn't really capture the student's true oral reading fluency?

These are issues of sampling. Sampling is extremely important with performance assessment. Because there are so many extraneous factors that can contribute to a student's performance on any single performance task, students need to complete a number of tasks to serve as the basis for a stable estimate of performance (Shavelson, Baxter, & Pine, 1992; Shavelson, Baxter, & Gao, 1993; Gao, 1996).

There aren't any hard and fast rules regarding sampling. A single, good culminating activity that relates directly to skills and products students have been developing all along may be enough. Or a simple, straightforward learning target, such as demonstrating how to use a band saw, might only require a single instance of performance.

However, we often need to gather several samples to get a stable view of student achievement because the behaviors we are assessing are complex. For example, to determine how good students are at solving math problems we would have students use various problem-solving strategies superimposed over a range of content. Or, to determine reading fluency, we might choose several passages from books with different content and purposes—narrative, expository, or persuasive text. As another example, using the writing assessments from the *Work Keys Assessment*, Gao (1996) found that six prompts and two raters were needed to substantially increase ability to generalize about student ability to write.

In situations outside the classroom, such as statewide testing, it can be very expensive and time consuming to collect enough samples of student performance to make stable judgments about individual students.

Enter the portfolio. Although Chapter 11 focuses on portfolios, it is appropriate to mention them here, as well. One way to increase the number of samples that show student progress and status is to have them collect work in a portfolio. If we carefully outline for students the types of entries to use, we can get a good sample. We might require that a writing portfolio have *X* number of samples of narrative, expository, and persuasive writing. Or, we might require that a math problem-solving portfolio have *X* samples of work demonstrating different problem-solving strategies using different content.

Several states have attempted to implement portfolios to supplement statewide testing. For example, in Oregon, teachers are supposed to make sure that students have a portfolio to corroborate evidence from the statewide assessment.

Performance assessments don't all have to be given at the end of instruction. We can gather evidence over time, as with a portfolio, to make the sample adequately reflect the breadth and depth of the learning target being assessed. In fact, the types of learning targets covered by performance assessment—reasoning, skills, and products—fit nicely into tracking progress over time using a portfolio.

Portfolios, however, are not just a way to increase the sample of student work so that we obtain more reliable evidence of student performance; they are also a way to involve students in their own assessment to assist them to take control of their learning. They offer so many benefits to student motivation and learning that we would never recommend using portfolios for the sole purpose of increasing sample size. Chapter 11 will offer more information regarding the student-involvement aspects of portfolios.

One final thought about sampling: As with other methods of assessment, we need to consider the importance of the decision to be made from the information to plan sample size. For important decisions, such as placement in a special program or graduation from high school, we need a bigger sample because the decisions can seriously affect a student's life and we want to make sure we are correct. But other decisions we can reconsider easily. For example, if we don't get an accurate picture of student ability to give an oral presentation due to some temporary interfering condition, such as the student being overtired, we may easily retest and replace the erroneous information.

TRY THIS

Activity 7.2 Determining Sample Size

Choose one of the learning targets you listed in Activity 7.1. Consider the breadth and depth of this target. How many samples of student performance might you need to be sure your inference about student proficiency is accurate? Write that number down and save it, noting how you might have to vary the tasks to cover the breadth and depth of the target.

Stage 2: Develop the Assessment—Rubrics

In the case of performance assessment, we need to develop both tasks and scoring guides. In this section we will examine what good rubrics look like and outline how to develop such rubrics.

Rubrics—Distinguishing the Gems from the Duds

Although there is no single “good” rubric, scoring guide, or performance criteria list, there are lots of bad ones out there. Simply type the word *rubric* into any search engine and you’ll be confronted with hundreds of hits. How do you select those that are strong? We have developed a rubric for rubrics, a “Metarubric” if you will, to answer that question. Here are the steps we completed to develop the Metarubric, and the results at each step. (To work through these steps for yourself to understand better where the metarubric came from, we recommend completing Segments 2–5 of the ETS video, *Student-Involved Performance Assessment*.)

We brainstormed what we want performance criteria (rubrics, scoring guides) to do for us in the classroom, because uses for the rubric influence design. Here is the list that we generated:

- Define quality for ourselves.
- Describe quality for students.
- Make judgments more objective, consistent, and accurate.
- Improve grading consistency.
- Guide instruction.
- Provide a common language.
- Promote descriptive feedback to students.
- Promote student self-assessment and goal setting.
- Describe quality to parents.
- Make expectations for students explicit.
- Eliminate bias.
- Focus teaching.
- Track student learning.

Notice that the list contains both assessment *for* and assessment *of* learning purposes. In the classroom we want rubrics that can help us with both.

Next, we identified the features that rubrics need to have to serve these purposes. We read what others had to say, made our own lists, and looked at hundreds of sample rubrics to identify everything of importance. Here is our list of what makes a good rubric:

- Focus on what is important.
- Be clear enough for everyone to understand.
- Define various levels of success.
- Be available in student-friendly language.
- Include only those aspects of a performance or product that are most valued.
- Include what is valued most as major parts of the rubric.
- Align with standards.
- Have a user-friendly format.
- Provide directions for use.
- Make language consistent across levels.
- Make levels distinguishable.
- Use no “fudge words,” such as *adequate* or *sometimes*.
- Have models to illustrate what is meant.
- Contain descriptive detail.
- Define terms.
- Have visuals to reinforce definitions.
- Use non-value-based adjectives.
- Don’t be negative at the low end.
- Be age appropriate.
- Match important goals.
- Include information on what the student did right at each level.
- Make it clear how to differentiate between score points.

Obviously, this list is way too long to be useful, so we looked for logical groupings of ideas and noticed that they seemed to fit nicely into four categories. We call these categories *traits*, and they are shown in Figure 7.3. Note that at this point in our rubric development, we have not yet described levels of quality for each trait. Figure 7.3 merely shows the traits with some definitions. The whole Metarubric, with scoring levels defined, can be found on the CD in the file, "Metarubric."

The Metarubric can look a bit daunting at first, so we will ease into it by just using the trait summary in Figure 7.3 to critique the oral presentation rubric in Table 7.3.

Content

Take a look at the content of the oral presentation rubric in Table 7.3 and answer the following question: "Does it cover everything of importance and does it leave out unimportant things?" (Remember, think analytically, and put aside any comments about clarity or practicality or fairness and focus just on the trait of *Content*. We will look at the other traits in a moment.)

On the positive side, the rubric includes much that is important—delivery, language, and organization. On the negative side, did you notice that there is no place to rate the quality of the oral presentation's content? You may also have noticed that aspects of good delivery and good language choice are missing. We would say this rubric is a little more than halfway there in terms of content.

Clarity

Analyze the clarity of the rubric in Table 7.3 using the Metarubric summary in Figure 7.3. What questions do you have? On the positive side, the various levels of the rubric are defined and are fairly descriptive, and there are samples of student performance that illustrate the various score points on each trait. On the negative side, people frequently ask about the score scale—why there are two scores in each range and how you differentiate between them. Others note words that students would not understand and point out the lack of definition of each trait. So, for the trait of *Clarity*, this rubric is about halfway there.

Practicality

Analyze the rubric in Table 7.3 on the trait of *Practicality* in Figure 7.3. How easy would it be to use? On the positive side, it would be easy for teachers and students to internalize the number of traits and number of score points if a few other details were cleared up, such as how to handle the two score points in each box. It would also be useful as a diagnostic tool

Figure 7.3 Metarubric Summary

The Metarubric contains criteria for judging the quality of rubrics—a rubric for rubrics. There are four traits: *Content*, *Clarity*, *Practicality*, and *Technical Quality/Fairness*. The Metarubric summary defines each trait. The complete Metarubric (found on the CD) defines three levels of quality for each trait.

1. **Content:** What counts? What users see is what you'll get.
 - Does it cover everything of importance—doesn't leave important things out?
 - Does it leave out unimportant things?
2. **Clarity:** Does everyone understand what is meant?
 - Are terms defined?
 - Are various levels of quality defined?
 - Are there samples of work to illustrate levels of quality?
3. **Practicality:** Is it easy to use by teachers and students?
 - Will students understand what is meant? Is there a student-friendly version?
 - Can students use it to self-assess and set specific goals?
 - Is the information provided useful for planning instruction?
 - Is the rubric manageable?
4. **Technical Quality/Fairness:** Is it reliable and valid?
 - Is it reliable? Will different raters give the same score?
 - Is it valid? Do the ratings actually represent what students can do?
 - Is it fair? Does the language adequately describe quality for all students? Are there racial, cultural, or gender biases?

to plan instruction, although additional detail in the traits would assist in planning specific lessons and in giving descriptive feedback. We judge the rubric to be about halfway home on this trait, as well.

Technical Quality/Fairness

Finally, examine the rubric in Table 7.3 using the Metarubric trait of *Technical Quality/Fairness* in Figure 7.3. It does include information about rater agreement, but there is no information about how the rubric functions with different groups of students, how easy it is to achieve high levels of rater agreement, or how it works to score oral presentations with nonnative speakers of English. Some features in the wording might disadvantage

Table 7.3 Sample Rubric for Oral Presentation

Score	Language	Delivery	Organization
A = 5	<p>Correct grammar and pronunciation are used.</p> <p>Word choice is interesting and appropriate.</p> <p>Unfamiliar terms are defined in the context of the speech.</p>	<p>The voice demonstrates control with few distractions.</p> <p>The presentation holds the listener's attention.</p> <p>The volume and rate are at acceptable levels.</p> <p>Eye contact with the audience is maintained. The message is organized.</p>	<p>The speaker sticks to the topic.</p> <p>The main points are developed.</p> <p>It is easy to summarize the content of the speech.</p>
B = 4 C = 3	<p>Correct grammar and pronunciation are used.</p> <p>Word choice is adequate and understandable.</p> <p>Unfamiliar terms are not explained in the context of the speech.</p> <p>There is a heavy reliance on the listener's prior knowledge.</p>	<p>The voice is generally under control.</p> <p>The speaker can be heard and understood.</p> <p>The speaker generally maintains eye contact with the audience.</p>	<p>The organization is understandable.</p> <p>Main points may be underdeveloped.</p> <p>The speaker may shift unexpectedly from one point to another, but the message remains comprehensible. The speech can be summarized.</p>
D = 2 F = 1	<p>Errors in grammar and pronunciation occur.</p> <p>Word choice lacks clarity.</p> <p>The speaker puts the responsibility for understanding on the listener.</p>	<p>The student's voice is poor.</p> <p>The volume may be too low and the rate too fast.</p> <p>There may be frequent pauses.</p> <p>Nonverbal behaviors tend to interfere with the message.</p>	<p>Ideas are listed without logical sequence.</p> <p>The relationships between ideas are not clear.</p> <p>The student strays from the stated topic.</p> <p>It is difficult to summarize the speech.</p>

Samples of student work illustrating levels of quality are available. Research information on technical quality: Exact agreement rate on scores is about 70%.

Source: Unknown.

certain groups of students. For example, consider the requirement of “correct pronunciation.” Correct for whom? What about nonnative speakers or those who speak a dialect of English? Consider also the phrase, “The message is organized.” Different cultures value different organizational patterns. What will we hold students accountable for? This rubric is about halfway home on the trait of *Technical Quality*.

Common Problems with Instructional Rubrics

In our experience, there are a few problems that crop up frequently in rubrics. Most of these problems relate to the Metarubric trait of *Content*. One relates to the trait of *Clarity*.

Counting, When Quality Is More Important Than Quantity

In an attempt to increase the objectivity of scoring guides, it is tempting to count things—the number of spelling errors, the number of references, the number of topics covered, and so on. But, most often this backfires: aren't 2 really good references better than 5 bad ones? Aren't spelling errors on 10 different words more of a problem than 10 spelling errors all on the same word? Couldn't a thorough coverage of a single topic be more impressive than cursory coverage of several topics?

Figure 7.4 shows a writing rubric based on counts.

What message does this rubric send to students about what is important in writing? How well does it define quality? What assistance does it provide on instructional decisions? If I, as a student, understood the rubric, might I be able to write a nonsensical essay and still get a good score?

The bottom line: For classroom uses, don't use counts to indicate quality unless there is compelling evidence that a count actually is the important criterion, that it does define quality.

Important Details Left Out

Let's say that I play the violin and that I'm in a contest. Let's further say that, although I feel my bowing is strong, bowing is not included in the criteria, which evaluate intonation, rhythm, and interpretation. What message does this send me about bowing? At the very least it sends the message that bowing is unimportant. But, more of a problem is that I get no feedback on an important part of my skill so that I know what improvements I need to make.

Three unfortunate things happen when we leave important things out of a rubric: (1) we send the message that what is left out is unimportant; (2) we generate incomplete infor-

Figure 7.4 Rubrics with Problems—Writing Rubric Based on Counts

<p>Introduction</p> <p>3 Introduces the topic and includes 4 to 5 sentences</p> <p>2 Introduces the topic and includes 2 to 3 sentences</p> <p>1 Does not relate to the topic, or includes 1 sentence</p> <p>Body Paragraph 1</p> <p>3 Has a topic sentence and at least 3 supporting details</p> <p>2 Has a topic sentence and 2 supporting details</p> <p>1 Has no topic sentence or only 1 supporting detail</p> <p>Body Paragraph 2</p> <p>3 Has a topic sentence and at least 3 supporting details</p> <p>2 Has a topic sentence and 2 supporting details</p> <p>1 Has no topic sentence or only 1 supporting detail</p> <p>Body Paragraph 3</p> <p>3 Has a topic sentence and at least 3 supporting details</p> <p>2 Has a topic sentence and 2 supporting details</p> <p>1 Has no topic sentence or only 1 supporting detail</p> <p>Conclusion</p> <p>3 Summarizes the main points and includes 4 to 5 sentences</p> <p>2 Summarizes the main points and includes 2 to 3 sentences</p> <p>1 Does not summarize the main points, or includes 1 sentence</p>
--

mation on which to plan future instruction; and (3) we provide no feedback to students on the quality of valued elements. For example, let's assume that a state leaves "voice" and "word choice" out of its rubric for writing; it includes only main idea, support, consistent focus, organization, sentence construction, and mechanics. Leaving out important aspects of quality might be done for the very best of reasons—state developers might be afraid that they will not get consistent scores on these traits. But, what do you think might happen to

writing instruction if such a rubric were used on the statewide assessment? What message does use of such a rubric send to students about the nature of quality writing? What diagnostic benefit is lost to teachers?

Making sure that the correct information is in rubrics is not just an intellectual exercise; it is personal for students. For example, read the student essay, "Junk Food," in Figure 7.5. The paper certainly has weaknesses. The ideas are incorrect, the organization rambles, grammar and spelling are weak, and sentences don't flow. But, there is one big strength—voice. You want to read the paper because the voice draws you in. What if this student lived in the state where voice was not scored as part of the state assessment? The student would receive the message that his writing has no redeemable value, when actually the student has one huge strength to build on—compelling voice. What impact might this omission have on the student? Similarly, the teacher would receive no feedback about the student's biggest strength, so she might not take it into consideration when planning instruction for this student.

The bottom line: Always include everything of importance on a rubric, even if it is difficult to define. The things most difficult to define are actually those that *most* need definitions, and developing a descriptive rubric is a fine way to proceed. Unless we can define important learning targets so that everyone knows what they mean, how will we ever be able to help students achieve those goals? If we leave important things out of a rubric, what incorrect messages might we send to students about the nature of quality?

Irrelevant Details in a Rubric

The flip side of ensuring that important features of performance are included in a rubric is ensuring that irrelevant details are left out. Imagine a scoring guide for a poster that illustrates knowledge of dinosaurs, for which one criterion is, "Must have three colors." Why does a poster have to have three colors? Can a black-and-white poster be highly effective in communicating information? What message does the requirement of three colors on a poster send to students about the nature of quality? What is our real reason for requiring three colors? If our goal is to teach students to create posters that capture people's attention and enhance the message, the criterion might be better phrased, "The design of the poster draws people in and supports the message." Then the teacher can show students how effective use of color can help accomplish this.

The bottom line: Go for importance. Of course importance implies that we understand the target well enough to know what *is* important. Developing a rubric for "presentation" is harder than checking off "three colors." But it is more useful in teaching students the true nature of quality and in planning instruction toward that end.

Figure 7.5 Junk Food (Student Essay)

Everything in the world has to have food may it be good food or junk food. Junck food is one of the more populare food. Most people like soda pops, hambriger, popcorn, shakes itc. Some of the places you can get these at is at 7 eleven stores, Mcdonalds, Dariy Queen etc.

Some Health food nut say that you will get fat if you eat hamburgere. You will but if you just at health food al the time, your body will get to meny vitimens and you can die. Health food is a food that will give your vitemans and cleans out yore iners.

Eny whay you nead junk food to get your adrental gland working. Junk food like hamburger is good food. Some people say that it is bad for you or is it. I don't think it is so bad for you because you get tomatos, lettice, musterd, relish and meat. Shakes. Shakes are made out of mile, ice, and aritvial flaver and suger.

Some people say that fried chicken is good for you but is it. the chicken is fried in nothing but oil. It is one of my favorit food so I don't care what eny one thinkes. If it is food it is food.

If you just had a candy bare out in the desert would you just throw it a way hec no. If you were out on the dessert you would take one bite of it and keep it in your mouth for days.

If you don't eat eny food you will die in 63 days of what is called starvation. Starvaison ocures wen you don't get enough food and you don't get eny oxegen to your bones.

Source: Anonymous student work.

Why Are We Using Rubrics Anyway?

What's the purpose of our rubrics—to enable students to get a good score or to teach them the nature of quality? If the goal is just to enable students to get a good score, we can put anything on a scoring guide: five references, three colors, fill the page with writing, three paragraphs on each topic, has a picture on the cover. But, if our goal is to define quality, then we need to be much more careful about what we include in our rubrics.

The bottom line: The goal of using scoring guides should be to define quality, and not just to provide a scoring mechanism or justification for giving a grade.

Student-Developed Rubrics Where Anything Goes

We are in favor of involving students in developing the criteria by which their work will be judged. Students can help us refine our thinking and help us spread out the work. Involving students helps them internalize the criteria and bring them to bear on their own work. But, we're also in favor of leading students toward good criteria—criteria that define quality. Most often, we, as teachers, know more about the criteria for quality than students do.

If students come up with criteria along the lines of “three colors” for a quality poster, we need to be prepared to broaden their thinking by showing them examples of effective posters with fewer (or more) than three colors. If they say that work has to be two pages long, we need to be ready to show them effective work that is only one page long, or that is six pages long.

The bottom line: Student involvement in developing criteria is a powerful strategy if done well, but it can lead to misunderstandings about the nature of quality if not done well. Anything does not go. Be prepared to assist your students in their discovery of good criteria through use of thoughtfully chosen examples.

Skimpy Scoring Guides

The previous problems have related to the Metarubric trait of *Content*. This problem relates to the Metarubric trait of *Clarity*. Have you ever seen a rubric like the one in Table 7.4? Sometimes the categories are more clearly fleshed out, but the details for how to assign the number of points are meager or nonexistent. We call these “skimpy rubrics” because there is little help with defining what is meant, or determining levels of quality. Will different teachers be likely to give the same scores to the same papers using this rubric? Will this help a single teacher be consistent in scoring over time and across assignments? Will this kind of rubric help students understand what their score means and will they be able to use the rubric to self-assess and set goals for improvement? We think not.

Compare the rubric in Table 7.4 to the “Central Kitsap Student-Friendly Math Scoring Guide” in the CD file, “Rubric Sampler.” Notice that the Central Kitsap rubric defines each trait and provides descriptors that help teachers and students find the correct level of performance. The scoring guide also is illustrated with student work at each level of quality at each grade level. Although imperfect, the Central Kitsap guide is definitely headed in the right direction.

The bottom line: Definitions, descriptive language, and examples are what make rubrics clear. Look for other skimpy and well-defined scoring guides in the CD file “Rubric Sampler.”

Table 7.4 Skimpy Mathematics Scoring Guide

	5	4	3	2	1
Understands the problem					
Problem solving					
Correct calculations					
Communication					

5 is high and 1 is low

Developing General Holistic and Analytical-Trait Rubrics

There are different types of rubrics. In Chapter 6, we recommended using *task-specific* scoring for extended written response exercises—that is, scoring guides developed for each individual task, assigning points to specific information contained in the response. If the intent is to assess knowledge, as is the case with many extended written response questions, then the scoring guide needs to reflect the specific information expected, and a task-specific rubric is appropriate.

We don’t recommend task-specific rubrics for the types of learning targets described in this chapter—reasoning, skills, and products. For a longer discussion of this point, please see Chapter 2 in *Scoring Rubrics in the Classroom*, the book accompanying ETS’s video, *Student-Involved Performance Assessment*. All we will say here is that if we want students to generalize from task to task what they are learning about quality problem solving, making inferences, giving oral presentations, working in a group, or designing experiments, then we need to develop and use general rubrics—those that can be used across tasks. “The influencing of instructional practices to date has been served most powerfully by generic rubrics” (Khatti, Reeve, & Adamson, 1997, p. 4).

Because we are most interested in the instructional aspects of rubrics, from this point on we will concentrate on how to develop general, not task-specific, rubrics.

What “counts” in a performance? What distinguishes levels of quality? What does it look like when a student has reached different levels of competency? What are the essen-

tial characteristics that define quality? In the following subsections we discuss the steps we use to clarify the nature of quality and to develop rubrics ourselves, using as our example first-year foreign language studies. As we go, consider how you might apply the process to develop a rubric for one of the learning targets you listed in Activity 7.1. These same steps can be used to develop rubrics for any product or performance, such as group collaboration, oral presentations, lab reports, self-reflection, reading fluency, critical thinking, art criticism, or writing.

Steps in Rubric Development

- Step 1* Establish a knowledge base.
- Step 2* Gather samples of student performance.
- Step 3* Sort student work by level of quality.
- Step 4* Cluster the reasons into traits.
- Step 5* Identify sample performances that illustrate each level.
- Step 6* Make it better.

Rubric Development Step 1: Establish a Knowledge Base

Determine your prior knowledge about quality and begin in one of the following ways. If you're an expert in creating the product or performance yourself, merely write down the features that distinguish levels of performance and refine them through looking at samples of student work. If you feel as if you know what quality work looks like when you see it, but you have difficulty writing it down off the top of your head, engage in the entire process given here.

Sometimes you have no idea where to begin. "What the heck is critical thinking?" "What in the world do they mean by 'lifelong' learner?" In this case, you must talk to others, examine state standards or curricula, and read what experts in the field have written. How do they define quality? Pursue these steps in conjunction with the following process.

In all cases, it is to your advantage to review as many existing rubrics as you can. Why recreate the wheel? Depending on your level of expertise, you can use sample rubrics ahead of time to help you understand what others mean, you can use sample rubrics to corroborate or extend your own thinking, or you can use them as you finish to help you refine your work.

Rubric Development Step 2: Gather Samples of Student Performance

Gather a range of student performances on the reasoning proficiency, skill, or product under consideration. For example, foreign language teachers wanted to develop a rubric for the following learning target: “comprehends and uses spoken language to satisfy social demands.” They wanted a rubric to use across languages—French, Spanish, German, and Japanese. So they audio- and videotaped students conversing in the target languages about everyday topics: their families, the area in which they grew up, the weather, their school experiences, and so forth.

Rubric Development Step 3: Sort Student Work by Level of Quality

At this step, sort the samples into three stacks according to level of quality—strong, middle, and weak. Ask yourself, “What makes the strong stack different from the middle stack? What makes it different from the weak stack?” Use the samples to create a list of descriptors of quality at various levels. We recommend trying to come up with as broad and long a list as possible. The foreign language teachers labeled their stacks *proficient*, *mid-range*, and *emergent* because those terms are more positive at the low end of achievement. As they sorted, they wrote down their reasons for placing each student performance in its chosen pile. They kept sorting until they were not adding anything new to their list. Table 7.5 shows what the foreign language teachers came up with.

Table 7.5 Foreign Language Teachers’ Brainstormed List of Features Defining Quality

Proficient	<ul style="list-style-type: none"> Bows and excuses self on approach and leaving (Japanese) Speaks in paragraphs Offers additional information Responds immediately and appropriately to both rehearsed and unrehearsed questions Combines several sentences Uses confirmational questions naturally Responds to novel questions/may ask for clarification Speech is approaching natural speed Initiates conversation Pronunciation is accurate Uses a full range of first-year grammar structures correctly Elaborates Uses complete sentences when appropriate Expands response in order to respond to implied questions Asks questions of the examiner Vocabulary is broad enough to sound natural
-------------------	---

Table 7.5 (Continued)

Mid-Range	<p>Bows and excuses self when approaching the teacher (Japanese), but not when leaving</p> <p>Combines learned phrases with interjectives</p> <p>Responds rapidly to rehearsed questions, but few unrehearsed questions</p> <p>Responds appropriately to rehearsed questions, but not to unrehearsed questions</p> <p>Pronunciation causes no confusion</p> <p>Verbs and endings are generally in place; the student can correct if necessary</p> <p>Relationals are generally in place; the student can correct if necessary</p> <p>Asks for clarification in the foreign language</p> <p>Sometimes repeats a question in English before answering</p> <p>Speech speed is slow, but causes no discomfort</p> <p>Speaks in full sentences, but not paragraphs</p> <p>Responds to questions, but doesn't ask any of her own</p> <p>Sometimes has to come up with awkward phraseology to compensate for a moderate vocabulary</p> <p>Anglicizes word order, but the message is communicated</p> <p>Gives desired information, but no more</p> <p>Self-corrects</p> <p>Frequently asks for rephrasing</p>
Emergent	<p>No sign of deference to the teacher (Japanese)</p> <p>Replies in single word utterances or fragmentary speech</p> <p>Responds inappropriately to rehearsed questions</p> <p>Speech is slow and causes confusion and discomfort</p> <p>Pronunciation is anglicized and causes confusion</p> <p>Intonation causes discomfort</p> <p>Hesitates before answering, causing discomfort</p> <p>Rehearsed expressions are confused or misapplied</p> <p>Verbs and endings are confused or misapplied</p> <p>Relationals are misapplied, causing confusion</p> <p>Vocabulary is limited</p> <p>Pronunciation confuses phonemes</p> <p>Misuses of grammar interfere with communication</p> <p>Uses random word order</p> <p>Hesitates excessively</p> <p>Understands few questions</p> <p>Needs frequent repetition and/or clarification of questions in English</p>

When creating your list, include as much detail as you can. Go beyond general terms typically used. Students may not understand “lacks fluency,” but they will understand “speaks slowly with hesitation.” To generate detail, ask yourself questions such as these: “What specific features made me judge that the speech lacks fluency?” “What am I saying to myself as I categorize the performance?” “What descriptive feedback might I give to this student?” If you want the rubric to provide descriptive feedback and to function as a self-assessment and goal-setting tool for students, it is to your advantage to include those descriptive phrases from the outset.

Rubric Development Step 4: Cluster the Descriptors into Traits

Your sorting and describing will result in a hodgepodge of descriptors at each level of performance. Some invariably will overlap or cluster together into the same category of performance. Invariably, someone will say, “Wait a minute. ‘Speaks in paragraphs’ is the same as ‘Combines several sentences. Why not delete one?’” Or, “Wait a minute. We have a whole lot of statements that refer to fluency. Why not group them together?” Or, “I had trouble placing a student performance in a single category because it was strong in fluency but weak in pronunciation. Let’s score performance on those two dimensions separately.” That’s exactly the purpose of Step 4—to determine what major categories, dimensions, or traits seem to be emerging from the sorting process, eliminating redundancies and including descriptors from the brainstormed list that best describe the traits at each achievement level.

The foreign language teachers identified four traits that seemed to cover everything in their lists: pronunciation, grammar, fluency and content, and social skills. They built a draft rubric by deleting redundant descriptors and sorting the remaining statements at each level of quality into these four traits. Table 7.6 shows their draft rubric.

During this process you might find that you need more than three levels of quality because you encounter examples that fall between levels. That is fine as long as you can find descriptors and/or sample performances that differentiate the levels. You can assign the numbers 5, 3, and 1 to the levels and allow 2s and 4s to represent those examples demonstrating characteristics of the two adjoining score points. As long as you get consistent scoring and you and students can differentiate performance levels, identify however many performance levels you need.

Rubric Development Step 5: Identify Sample Performances That Illustrate Each Level

Return to the samples categorized as strong, middle, and weak and select examples that illustrate well what is meant by each trait at each achievement level. These samples—also

Table 7.6 Draft Foreign Language Rubric

Score	Pronunciation	Grammar	Fluency and Content	Social Skills
3	Sounds natural Causes no confusion Accurate Accurate intonation	Verbs, endings, and relationals are accurate Uses the full range of first-year structures	Speaks in paragraphs Offers additional information Response is immediate and appropriate Responds to novel questions without hesitation Speech approaches normal speed Initiates conversation	Bows and excuses self on approach and leaving
2	Beginning to sound accurate; doesn't interfere with communication Can self-correct	Verbs, endings, and relationals are sometimes correct and sometimes incorrect Can self-correct if requested Word order is anglicized but the message is communicated Only simple structures are used well	Tends to hesitate, but responds appropriately to most rehearsed questions Responds rapidly to rehearsed questions, but hesitates with novel questions Tends to repeat the question in English before answering Asks for clarification in foreign language Gives the desired information, no more	Bows and excuses self only when approaching or leaving the teacher, but not both
1	Mostly anglicized Phonemes are confused Interferes with communication Causes discomfort Can't self correct	Uses random word order Verbs, endings, and relationals are mostly confused or misapplied Grammar problems cause confusion in communication Can't self-correct even with prompting Not even the simplest structures are used correctly	Uses single-word utterances Tends to respond inappropriately to rehearsed questions Speech is slow and causes confusion and discomfort Hesitates a lot Rehearsed expressions are confused or misapplied	Does not show deference to the teacher

called *models*, *exemplars*, *examples*, and *anchors*—help teachers attain consistency with each other and within their own scoring across time, students, and assignments. Samples also help students understand what each achievement level looks like in concrete terms. As an added benefit, identifying the best performances to illustrate each trait and level can suggest refinements to the rubric.

Be sure to have more than one sample to illustrate each level. If you show students only one example of good performance, all performances might come out looking the same. Rather, show several performances that illustrate each level and trait.

Rubric Development Step 6: Make It Better

Rubrics are always works in progress. As you use rubrics, all of the following can happen:

- Traits are added, deleted or merged. For example, the foreign language teachers noticed that the trait of *Fluency* also included accuracy of the information presented, so they considered splitting *Fluency* into two traits: *Fluency* and *Accurate Content*. They ultimately settled on a single trait under the collective term, *Fluency and Content*.
- Traits and descriptors are refined. For example, in the Six-Trait Model writing rubric (see the CD file, “Rubric Sampler”) the trait *Sentence Fluency* was once called *Sentence Correctness*. To be strong on this trait, students had to use complete sentences. Then raters discovered that some good writing had to be scored lower because it contained incomplete sentences, even though the sentence fragments were used to good effect in enhancing the meaning. So the meaning of the trait was changed. What we really want to accomplish with students is not only complete sentences, but control of sentence structure—complete sentences or fragments used consciously to create a feel to the writing, or to support the meaning in the most effective way, given the audience, purpose and format for the writing. So the trait has become *Sentence Fluency*.
- The number of levels increases as raters feel the need to differentiate levels of quality further. Or, the number decreases as raters realize they cannot distinguish accurately between two levels.
- Developers notice that some characteristics are included at one level of the rubric but are not represented, either as a strength or a weakness, at other levels, so they revise the rubric to achieve parallel content.

- A state rubric that works well for the purpose of collecting achievement status information on students may not be as useful for classroom use because it is too skimpy—it does not have enough description to define the levels for students. In this case, teachers might make a descriptive, analytical-trait classroom rubric that parallels the state rubric. For example, the writing rubric in one state has only two traits: *Conventions* to evaluate capitalization, usage, punctuation, and spelling, and *Content, Organization, and Style*, which includes everything else—focus, details, organization, rhetorical style, and so on. At the classroom level, many teachers clarify the “everything else” by using the Six-Trait Model, which covers the same characteristics of good writing as does the state rubric, but is written to meet teachers’ and students’ needs for more detailed information. Improving student performance through use of the Six-Trait Model therefore directly transfers to higher performance on the state assessment.

Activity 7.3 gives you the opportunity to develop your own rubric using the steps given in this chapter.

TRY THIS

Activity 7.3 Develop a Rubric

Select one of the learning targets you identified in Activity 7.1 as high priority for performance assessment. Develop a general rubric for that target by looking at the rubrics of others, sorting student work, and devising a draft, following the steps given in this chapter. You can find more detailed instructions for rubric development in the CD file, “Rubric Development.”

To reduce the amount of time this takes, try the following:

- Work with other teachers on a rubric that everyone can use. It might be a lot of work at first, but once it’s finished, you have it forever.
- Get students to help.
- Save your scarce development time for those learning targets that truly need definition. Select the ones you are most fuzzy about, ones that many students have trouble understanding, or ones that state assessment data indicate need attention.

Stage 3: Critique the Assessment—Rubrics

With a performance assessment, we critique the rubric and the task separately. At this point in performance assessment development, we focus on determining the quality of the rubric by using the Metarubric, summarized in Figure 7.3 and found in its entirety in the CD file, “Metarubric.” Activity 7.4 provides practice in analyzing your own rubrics for quality with the Metarubric.

DEEPEN UNDERSTANDING

Activity 7.4 Analyze Your Own Rubrics

Choose your favorite scoring guides. Analyze them using either the Metarubric summary in Figure 7.3 or the full Metarubric on the CD in the file, “Metarubric.” Analyze them for each Metarubric trait. What do you conclude about how well your rubric scores on the four traits? If you are working with a learning team, you may want to discuss observations and conclusions at your next meeting.

For additional practice using the Metarubric to identify rubrics useful for instruction in the classroom, work through Segments 3 and 4 on the video, *Student-Involved Performance Assessment*, or turn to Chapter 4 in the book packaged with the video, *Scoring Rubrics in the Classroom*. You also may wish to critique the rubrics in the CD file, “Rubric Sampler.”

Stage 2: Develop the Assessment—Performance Tasks

Now we turn to developing performance tasks. In this section, as in the section on developing rubrics, we will explore what good tasks look like and describe how to develop tasks that accomplish what they are intended to.

Dimensions of Good Tasks

Although much of the emphasis in performance assessment is on performance criteria, the quality of the task must receive equal attention. The purpose of the task is to elicit the correct behavior on the part of the student so that it can be assessed by the scoring guide.

Thus, there must be alignment between the task and the rubric, and both must reflect the learning target accurately.

Here's an example of a time when the task did *not* elicit the correct performance. A number of years ago, a group of assessment developers were creating fifth-grade prompts for a statewide writing assessment. One they wrote to assess narrative writing went something like this: "You have all heard stories of how the elephant got its trunk or how the camel got its hump. Think of something in nature and make up a story about how it got that way." Many students produced not narrative, but expository writing—why elephants have trunks and why camels have humps. The prompt did not elicit narrative writing, so they couldn't evaluate the students' level of achievement in narrative writing. The tricky part with performance tasks is that if we don't elicit the right performance we can't use the results (if we can even assess the performance at all) as evidence of level of achievement.

To avoid this problem as well as others, we have developed a rubric for judging the quality of performance tasks. Using the rubric development process as outlined previously, we identified five traits: *Content*, *Clarity*, *Feasibility*, *Fairness and Accuracy*, and *Sampling*. These traits are summarized in Figure 7.6.

Content

The job of the performance task is to elicit the right performance from students so that proficiency can be judged accurately. Thus, the performance task, the performance criteria, and the student learning targets to be assessed must match up. If the learning target to be assessed is "Understands how the past affects our private lives and society in general" (Kendall & Marzano, 1997, p. 114), then the task must call forth a product or performance that directly demonstrates that understanding.

Our definition of sound content in a performance task differs somewhat from those you may find in other task development guidelines. Some tend to emphasize authenticity and stress such criteria as *open*, *active*, *collaborative*, *fosters persistence*, *challenging*, and *rich*. While these might, indeed, be relevant for many of the tasks we design to assess complex targets, they are not criteria for all performance tasks. It is possible to have simpler performance tasks that assess simpler targets. Consider, for example, reading rate. The task is very simple: Ask students to read aloud and calculate the number of words read per minute. *Open*, *collaborative*, and *rich* are not task requirements in this case. Indeed, were they present, they could compromise the results. Our rule of thumb is this: Simple target, simple task; complex target, complex task.

Figure 7.6 Task Rubric Summary

Content: What are students asked to do?

- The task elicits the right performance; it fits the target and performance criteria.
- Simple target → simple task; complex target → complex task.
- Worth the time spent on it; students will learn from the task.
- The amount of scaffolding supports the task without compromising it.
- Task is engaging/interesting to students.

Clarity: How clear are the instructions?

- Instructions tell students clearly what to do.
- Students are reminded of the performance criteria.

Feasibility: How practical is the task?

- Students have enough time.
- Proper materials and equipment are available.
- Rating can be accomplished within the time allowed.

Fairness and Accuracy: How fair is the task?

- There is nothing in the task that will give an inaccurate picture of student achievement.
- All students have a chance to shine.

Sampling: How well does the task represent the breadth and depth of the target being assessed?

- The task(s) adequately cover all dimensions of the learning target being assessed.

OR

- The task is part of a larger plan to cover all relevant dimensions over time.

There are just enough tasks to show students' level of proficiency; not too few, not too many.

The important consideration is that the task elicit the intended performance so we can observe and evaluate it. Since many targets evaluated with a performance assessment are complex, performance tasks tend to be complex, open, active, and challenging. However, they don't have to have all these characteristics to be sound.

Clarity

Tasks are stated in such a way that students know exactly what to do. A good task also includes the criteria that will be used to evaluate the quality of the performance or product. This helps students focus on the most important aspects to include.

Feasibility

The task can be completed successfully given the time and materials available in the classroom. It is "do-able." Additionally, *Feasibility* refers to scoring—are adequate time and resources available to observe performances and/or evaluate the quality of products?

Fairness and Accuracy

There is nothing in the task that will compromise students' ability to show what they can do. All students have an equal chance to shine. Examples of threats to fairness include the following:

- Exercises and tasks using a context (sewing, baseball) not familiar to some students
- Extensive reading or writing requirements when the achievement to be demonstrated does not relate to reading or writing learning targets
- Exercises or tasks incorporating components irrelevant to the target(s) to be assessed; e.g., oral presentation or group discussion components when the learning target is not oral presentation or group discussion skills
- Anything else that might interfere with students accurately demonstrating their levels of achievement; e.g., successful completion of the task relies on personality traits or cultural background

Sampling

The task or tasks adequately cover all dimensions of the learning target to be assessed, as described in the previous section in this chapter on assessment planning.

Plan the Performance Task

To develop a performance assessment task, we follow the same procedure used in developing an extended written response exercise as described in Chapter 6:

- Specify the learning to be demonstrated.
- Specify the materials and constraints within which achievement is to be demonstrated.
- Remind students of the criteria that will be used to evaluate their performance or product.

Here is an example: “Using a stopwatch and a measuring tape, you are to use your knowledge of the physics of motion to determine the percentage of vehicles that exceed the speed limit as they pass the school. Then you are to write a report in which you explain your experimental design and share your results. Your report will be scored on experimental design, understanding of physics equations, and collection and presentation of results. Rubrics are attached.”

Tasks can vary along several key dimensions. Think carefully about the following options as you select or create tasks.

One Right Answer or More than One?

Here’s a math problem that has one correct answer: “A group of 8 people are all going camping for 3 days and need to carry their own water. They read in a guidebook that 12.5 liters are needed for a party of 5 people for 1 day. Based on this guidebook, what is the minimum amount of water the 8 people should carry?” Here’s a sample math problem with more than one right answer: “Estimate the amount of television advertising the typical American fifth grader will see in a year.” Both are good tasks; which option you choose depends on which is better suited to the learning target. Open-ended tasks (those with more than one right answer) often encourage more diversity in solution strategies, but the “how much water” problem, a one-right-answer problem, also lends itself to a number of possible solution strategies.

Written, Oral, or Visual Instructions, Activities, and Responses?

Selecting from among these options requires consideration of who and what you are assessing. When demonstrating a performance skill, students will be performing in front of one or more raters. However, the instructions can be written, oral, or visual, depending on students’ needs.

When assessing writing in English, the response will, of course, be a written product. But again, the instructions can be written or oral. When assessing young children's ability to construct a story, the instructions will be oral and the product may very well be visual—a series of pictures that tell a story.

Reasoning is usually assessed by means of a written product, but if student needs or other context factors so dictate, you can have students describe their reasoning orally.

Choice or No Choice?

This is a trickier one. Student choice can increase student motivation. Yet, giving students a choice of tasks to perform only makes sense if all the options are of equal level of difficulty and each adequately samples performance for the intended learning target. If, instead of completing several tasks, students are allowed to choose one, they may select the topic with which they are most familiar, in which case you will be unable to determine their performance across the range of tasks that represent the target in question. Or, students might choose to respond only to narrative writing prompts and not to persuasive writing prompts and you will be unable to obtain a good estimate of ability to write for a variety of purposes. Or, students might choose to read aloud only passages with which they are already familiar and you will not get a good estimate of reading fluency.

For these reasons, standardized tests rarely offer students a choice of tasks to perform. There is more leeway in the classroom. Give students no choice if you need to determine whether they can do a specific thing. Give students choice if you will be giving them several opportunities to perform on a range of topics of similar difficulty.

The answer is different if all students are given the same task and they can choose how to respond. For example, if the target is a reasoning proficiency, students might demonstrate their reasoning through writing, an oral presentation, or a visual display.

Individual Work or Group Work?

In an attempt to make assessments as realistic as possible, teachers often have students work in a group. This is fine if one of your targets is to determine how well students work in a group. But, if your target is to ascertain individual achievement, a group product or presentation will not suffice. To get around this problem, some test developers have students work in groups to make sure everyone is beginning the task with a similar knowledge base. Then students work alone to create the product or perform the skill. In general, though, when assessing individual student reasoning proficiency, skill level, or product development capabilities, the task must be completed by each individual.

Spontaneously Occurring or Separate Event?

In a spontaneously occurring event, you evaluate the learning in the course of regular activities. For example, you may look for students' mastery of subject-verb agreement in writing done for other purposes. Sometimes, however, a separate, planned assignment is easier to control. It might take a long time to determine whether individual students can use subject and object pronouns correctly in their writing without an assignment designed to call forth evidence of that learning target.

Timed or Untimed?

Ideally, you would give students as long as they need to complete a task, but obviously there are some upper limits. You can allocate a certain amount of class time to get started and then students can choose to spend as much time completing the task as needed outside of school. Or, you can specify enough time so that almost all students will finish with no problem, and then offer options as needed for the remaining few who don't.

Figure 7.7 represents a worksheet for planning performance tasks. It is also on the CD in the file, "Performance Task Plan."

Figure 7.7 Plan for a Performance Assessment Task

<p><i>What learning target(s) will I assess?</i></p> <p><i>What will students do?</i></p> <ul style="list-style-type: none">• What knowledge are they to use?• What are they to perform or create?• What conditions are they to adhere to?• How much time will they have? <p><i>How many tasks will I need to sample well? How should these tasks differ to cover the breadth and depth of what I am assessing?</i></p> <p><i>By what criteria will I judge the performance or product?</i></p>
--

TRY THIS**Activity 7.5 Develop a Performance Task**

Using Figure 7.7 as a guide, develop a performance task for one of the learning targets you listed in Activity 7.1. If you are working as a part of a learning team, share your task with the team and ask them to critique it using either the Task Rubric summary in Figure 7.6 or the complete task rubric found in the CD file, “Performance Task Rubric.” Note any suggestions for revision.

Creating Tasks to Elicit Good Writing

Writer Donald Murray tells us that “a principal cause of poor writing received from students is the assignment . . . they have to be well prepared so that the students know the purpose of the assignment and how to fulfill it. Far too many teachers blame the students for poor writing when the fault lies with the teacher’s instructions—or lack of instructions” (Murray, 2004, pp. 98). If the goal of the task is to create a written product, it can be helpful to answer the questions that writers must ask to write well:

- What is my role?
- Who is my audience?
- What is the format?
- What is the topic?
- What is the purpose?

These questions are represented by the acronym RAFTS and are illustrated in Figure 7.8. Figure 7.9 shows how we might use these questions to plan the ingredients for a written task in a content area.

Role

Sometimes we ask students simply to write as themselves, as students, but we can often increase their motivation and the relevance of a task if we ask them to assume a role. Think about these questions to devise student roles: Who might be writing about this topic? If it is

a content-area topic (social studies, science, mathematics, health, art, and so on), who in the practice of this content area might be writing about this topic? What job might they have?

Audience

When we don't specify an audience, students are writing to us, their teachers, by default. Yet they don't often know how to envision their audience's needs when they write, even if it is us. For writers to make good decisions about what information to include, what terminology to use, and what tone to adopt, they need to think about to whom they are writing. When we ask students to write thorough explanations, it is helpful if we specify an audience who is not familiar with the topic. If we are the audience, either stated or unstated, students often conclude that we know plenty about the topic already. It is hard to write about something to someone who knows more about it than you do, and that particular circumstance doesn't occur in life beyond school very often. In life beyond school, when we are writing to inform, generally, the audience doesn't have the same level of expertise as the writer does. So, in tasks calling for informational writing, consider specifying an audience who might need to know the information and who doesn't already know it.

Figure 7.8 RAFTS Task Design

<p>Role (Writers must imagine themselves as fulfilling specific roles—for example, as tour guides or scientists or critics—when they write.)</p> <p>Audience (Writers must always visualize their audiences clearly and consistently throughout the writing process. If they don't, the writing will fail.)</p> <p>Format (Writers must see clearly the format that the finished writing should have, whether brochure, memo, letter to the editor, or article in a magazine.)</p> <p>Topic (Writers have to select and narrow their topics to manageable proportions, given their audiences and formats.)</p> <p>Strong verb (words like “cajole,” “tempt,” “discourage,” when serving as definers of the predominant tone of a piece of writing, will guide writers in innumerable choices of words.)</p>
--

Source: From “Why Grade Student Writing?” by E. Smith, 1990, *Washington English Journal*, 13(1), p. 26. Reprinted by permission.

Figure 7.9 Content-Area Writing Planning Form

Each of the content areas we teach has beyond-school applications. Specifically, people use the knowledge and skills while working, engaging in hobbies, and performing everyday living tasks.

Think of one subject area you teach. Then think of a few jobs/careers that require extensive knowledge and skill in your subject area. Fill in the chart to help you think about “authentic” applications of writing in your field.

My subject area: _____

What job/career requires use of this subject? (ROLE)	What might this person be writing about? (TOPIC)	To accomplish what purpose? (STRONG VERB)	Who might be the audience for this writing? (AUDIENCE)	What format would the writing take? (FORMAT)

Format

This is a simple component of the task. If a beyond-school application of the writing would take the form of a report or an essay, then by all means specify that format. Decisions about format are driven by three considerations: audience, topic, and purpose. If our audience is primary students, our topic is insects, and our purpose is to inform, a report may not be the best format. We can convey interesting and important information about insects in an alphabet book, on a poster, in a sticker book, or on playing cards. If, on the other hand, our audience is a politician, our topic is water quality (as measured by the number and diversity of stream bugs found in water samples), and our purpose is to persuade the politician to take an action, a combination of a letter and a report will be more suited to the task.

Topic

We rarely leave this out of a task or assignment. However, even this aspect can cause student writing to be better or worse, depending on how it is crafted. When we specify the topic for students, we must exercise caution in how we state it. The question here is, are

we going to narrow the topic for students or are we going to expect students to narrow it for themselves? If we have been studying the foundations of the U.S. economic system and we want students to write about the Industrial Revolution, we will have to narrow the topic considerably for them to handle it successfully. Or, we can teach students how to narrow topics and let them determine the aspect they will focus on. Considerations in narrowing topics include who the audience is and how much time the writer will have. Generally, the less time, the narrower the topic. We can write all about the topic of friendship if we have a year or two and want to produce a book, but if we have only a week or so, we may wish to write a simple set of instructions for how to be a friend.

Strong Verb

This does not refer to strong verbs in the students' writing. Rather, in this context, *strong verb* refers to the verb we use in the task itself. What is the *purpose* for the writing? Most often writing tasks in school are set to accomplish one of three purposes—to narrate, to inform, or to persuade—and the forms of writing produced are often referred to as narrative, expository, and persuasive. In narrative writing, the purpose is to tell a story, either real (personal narrative or anecdote) or imagined (fictional narrative). In expository writing, the controlling purpose is to inform. In persuasive writing, we may have one of four purposes: to initiate thought, to change thought, to initiate action, or to change action. We may use both narrative and expository writing in service of persuasion, but the ultimate purpose for the writing is to cause something different to happen. Table 7.7 gives examples of verbs that help students understand what kind of writing they are to produce.

Table 7.7 Verbs Matched to Purpose for Writing

Purpose	Sample Verbs and Phrases	
To narrate	Describe an experience Entertain	Tell the story of Tell about a time when
To inform	Clarify Compare Define Discuss	Describe Explain Inform Teach
To persuade	Challenge Convince Defend	Incite Justify Persuade

So, what does it look like when these ingredients come together? Here is a short assignment (in which the student's role is simply to be an informed student): "Explain the mathematics formula we studied today in a memo to a student who was absent." Here is another short assignment: "Teach younger students how to read a contour map by creating a list of instructions accompanied by diagrams and/or illustrations." Figure 7.10 shows an example of a longer assignment following the RAFTS planning design. (Note that we use the word *purpose* in place of *strong verb* when calling out these elements for students.)

Figure 7.10 Example of RAFTS Task

To Your Health

Imagine that a fifth-grade teacher from the elementary school you attended has asked for your help. She is worried that her students don't understand how healthy childhood habits lead to becoming healthy adults. Because she knows that younger children look up to teenagers, she has asked you to teach her students about how to become healthy adults.

Your assignment:

In a report to be read by fifth-graders, explain the habits they can establish now to help them become healthy adults.

In framing your report, consider the following questions:

- What are healthy childhood habits?
- What does good health involve beyond healthy eating habits?
- What should a child do, and what should a child avoid?

Role	Well-informed older student
Audience	Fifth-graders
Format	Report
Topic	Health habits
Purpose	To teach (inform)

Your report will be judged on the basis of the attached criteria.

Stage 3: Critique the Assessment—Performance Tasks

At this point in performance assessment development, we focus on determining the quality of the task by using the rubric for tasks, summarized in Figure 7.4 and found in its entirety in the CD file, “Performance Task Rubric.” The CD file, “Performance Task Sampler,” contains sample performance tasks to critique. Activity 7.6 provides practice using the rubric to analyze your own tasks for quality.

DEEPEN UNDERSTANDING

Activity 7.6 Critique Performance Tasks

Gather one or more performance tasks you have used recently with students. Critique them for quality using either the Task Rubric Summary in Figure 7.6 or the whole rubric on the CD in the file, “Performance Task Rubric.” Note the revisions you think they need.

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

As we noted before, problems can still crop up in assessments, even with the best planning, so it is a good idea to keep notes on any facets of a performance task or rubric that may compromise students’ ability to show what they know and can do. Figure 7.11 lists potential sources of mismeasurement peculiar to performance assessment. If something appears to have gone awry, and you can’t identify the problem from this list, you may want to use the Performance Task Rubric and the Metarubric to troubleshoot your performance assessment.

Figure 7.11 Potential Sources of Bias and Distortion with Performance Assessment

- Lack of reading skills
- Inappropriate or nonexistent scoring criteria
- Evaluator untrained in applying scoring criteria
- Bias due to stereotypic thinking
- Insufficient time or patience to observe and score carefully
- Student doesn't feel safe
- Unfocused or unclear tasks
- Tasks that don't elicit the correct performance
- Biased tasks
- Students don't know the criteria by which they'll be judged

Seven Strategies for Using Rubrics as Instructional Tools in the Classroom

You're all set with a great scoring guide/rubric for evaluating a skill or product in your discipline. But how do you get students to understand and internalize your standards of quality? Performance assessment is a prime context for using assessment to help students learn. The instructional power here resides in using high-quality performance criteria to help students answer the three questions introduced in Chapter 2 to define assessment *for* learning: "Where am I going?"; "Where am I now?"; and "How can I close the gap?" Try out these seven strategies for using a scoring guide as a teaching tool and watch as students become competent, confident self-assessors and improve their performance in any subject.

DEEPEN UNDERSTANDING

Activity 7.7 Watch DVD Segment “Teachers on Rubrics”

Think back to rubrics you have used with students and the effects they have had on student learning. Then go to the accompanying DVD segment, “Teachers on Rubrics” for clips of teachers discussing the use of various rubrics in the classroom.

Where Am I Going?

Strategy 1: Provide a Clear and Understandable Vision of the Learning Target

Student-Friendly Versions of Rubrics

Rubrics are a great way to offer students a clear and understandable vision of the learning target. When we are using a rubric to define or clarify the target, sharing it with students at the outset makes sense. Many of our rubrics, unfortunately, are not yet written in language that makes sense to most of our students, in which case we will have to do a little work before sharing them. For example, a rubric to evaluate the characteristic “self-directed learner” designed for use with middle school students has the following statement in the “Accomplished” column for the first trait: “Takes pride in own work: constantly strives for excellence, works hard, aims to exceed standards, and fulfills all commitments in a timely manner.”

What is meant by all these words? We used the process from Chapter 3 to clarify the terms. First we identified the words to be defined, looked them up in a dictionary, wrote out their formal definitions, and then converted the formal definitions into student-friendly terms.

Words to be defined: *Exceed standards*—to go beyond a specified level of quality

Student-friendly language: I work to exceed standards. This means I understand what good quality work looks like and I try hard to make sure all my work is at least that good.

Words to be defined: *Fulfill commitments*—to pledge to a certain course of action

Student-friendly language: I fulfill commitments in a timely manner. This means I can keep track of what I have promised to do and make sure it all gets done on time.

TRY THIS**Activity 7.8 Student-Friendly Rubric Language**

Look again at the foreign language draft rubric in Table 7.6. Identify two words that students might not understand. Look these up in a dictionary and covert the definitions into student-friendly language.

Next, look at the rubrics you use in your classroom. Identify one to convert into student-friendly language. Rework the rubric so that students would understand the terms, using the procedures given here.

Introducing the Rubric to Students

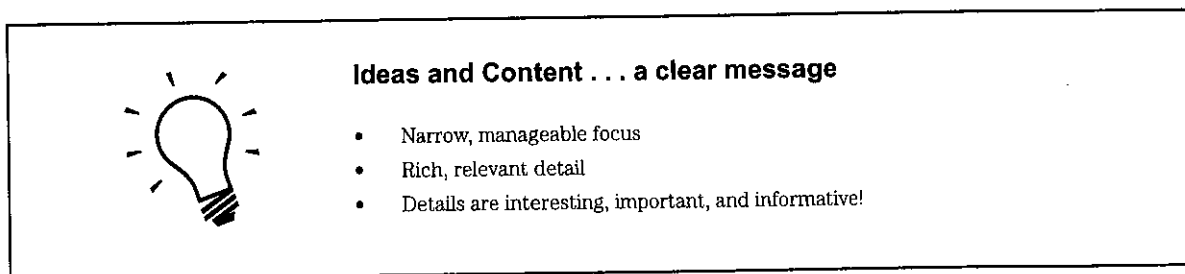
Before we hand a rubric out, we may want to introduce the concept of *quality* to students, so they have a better understanding of what they will be seeing in the rubric. In Chapter 2 we described the process we would follow in general terms. Here we use a specific example, illustrating how we would introduce a rubric for writing:

1. Ask students what is important in good writing and record whatever they say on chart paper. Don't try to get them to say certain words or phrases, and don't reword their responses. At this initial stage, we want to hear the words they use to describe quality as they understand it.
2. Read aloud a sample of writing having strengths they may not have thought of. (It can also have weaknesses, but the initial sample should be strong overall. We tend to choose a piece written by a student a little older than the group, selecting an example that will illustrate strong *Content* and *Voice*, two traits students often don't mention right away.) Ask students to think about their list of what is important in good writing. What might they like to add based on the piece they are listening to? You can repeat this with several samples over time and create a multipage list, or you can proceed to the next step after sharing one sample.
3. Tell students, "You have come up with a good list. Many of the things you have identified are also on the list that teachers and strong writers have created. Their list was also long—too long to remember everything—so they decided to see if they could group ideas into categories. They came up with X categories [where X equals the number of

traits in your rubric], which we call traits.” Then share the names of the traits in your rubric. If for example, you are using a six-trait writing rubric, you would show a poster or overhead transparency with the six traits listed.

4. Next, share your own definitions of the traits in your scoring guide. (We like to use bulleted lists for each trait, with phrases in student-friendly language as much as possible. Figure 7.12 shows a poster we would use to introduce the trait of *Ideas and Content* to upper elementary and middle school students.) Share definitions one trait at a time and ask students to identify similarities between their list and those you are sharing. “Did we say something about *Ideas and Content* on our list? Where?”
5. Pass out copies of your student-friendly scoring guide. Let students know that as a class, they will be using this scoring guide to examine and assess writing of all kinds—first the work of others, like the samples you have used, and then their own. (A student-friendly version of the six-trait writing scoring guide is on the CD in the file, “Rubric Sampler.”)

Figure 7.12 Sample Poster Introducing the Trait of Ideas and Content



Strategy 2: Use Examples and Models of Strong and Weak Work

Teaching Students to Use the Rubric to Evaluate Examples

First, gather models of strong and weak work—anonymous strong and weak student work, published strong (and weak, if available) work, and your own work. Share anonymous student samples that model both good work and problems students commonly experience, especially the perennial, pervasive problems. Here is an application of this strategy specific to mathematics. (Strong and weak samples of student work are on the CD in the files, “Samples of Student Writing” and “Student Math Problem Solving.”)

1. Choose one trait to focus on at a time. Find a mathematics problem that has anchor papers for each score point on your rubric.
2. Ask students to solve the problem. If you are working with younger students, you may wish to guide students through it using the following procedure:
 - Read the problem together. What is it asking you to do?
 - Underline important information in the problem.
 - Think about how you could solve this problem. What strategies could you use? What procedure will you follow?
 - As you solve the problem, show and tell your thinking. Show with pictures, charts, graphs, or diagrams. Tell using mathematical language to describe your reasoning, the strategies you used, and the procedure you followed.
3. Show an overhead transparency of a strong anonymous response to the problem. Have students score the response (don't tell them it's a strong response), for one trait using the student-friendly scoring guide. To do this (using a five-point scoring guide as the example), ask students to first decide independently whether the response is strong or weak. If they think it's a strong response, they read the scoring guide description of a "5;" if they think it's a weak response, they read the description of a "1." If the extreme ("5" or "1") description doesn't fit the response, students read the "3" description. If the response has some of "1" and some of "3," it's a "2." If it has some of "5" and some of "3," it's a "4." You may ask students to underline the statements in the scoring guide that describe the work they are examining.
4. Once students have settled on a score independently, have them talk in small groups to share their scores, using the language of the scoring guide to explain their reasoning.
5. Ask the class to vote and tally their scores on an overhead transparency. Then ask for volunteers to share their scores and their reasons. Listen for, and encourage, use of the language of the scoring guide.
6. Repeat this process with a weak anonymous sample student response, focusing on the same trait. Do this several times, mixing up strong and weak papers, until students are able to distinguish between strong and weak and give rationales reflecting the concepts in the scoring guide.

Thoughts on Strategies 1 and 2

Remember, when we say “score” in the context of formative assessment, we do not mean “giving a grade for the gradebook.” We mean evaluating anonymous student work to help students differentiate levels of quality. No grade is needed for this.

TRY THIS

Activity 7.9 Adapt Strategies 1 and 2

Think about how Strategies 1 and 2 would play out in your classroom. Translate their steps into one or more lessons you could use with your students to refine their vision of quality. Try the lessons out with your class. If you are working with a learning team, share what you did and how it worked.

Where Am I Now?

Strategy 3: Offer Regular Descriptive Feedback

Descriptive feedback points out to students their work’s strengths and weaknesses before it is too late—before the final grade—and models the kind of thinking we want them to do themselves about their work. If students have become familiar with the language of the rubric, we can use that language as the basis for descriptive feedback. If we are focusing on one trait at a time, we only need give descriptive feedback on that one trait. This has the effect of narrowing the scope of work for both the teacher and the student. With struggling students, we can show them they do indeed know some things and we can limit the things they need to work on at one time to a less daunting, more manageable number. Our feedback may be verbal, such as that given in a brief student-teacher conference, or we may choose to offer written feedback.¹

The Three-Minute Conference

If you confer with students as a way to offer feedback on their work, consider asking them to do some thinking prior to meeting with you. This causes the conference to take less time and your feedback to be more meaningful. For this activity, you can have students complete the form in Figure 7.13 (also available on the CD in the file, “Using Feedback to Set Goals”).

1. Identify a focus for the feedback—narrow it, if needed. Have them focus only on a few aspects of quality—either you choose the aspects of quality based on what you have been teaching them to do or let them choose, depending on their level of sophistication. (For example, in writing, a teacher may be focusing on how to include details that are interesting, important, and informative, which is part of the trait of *Ideas and Content* in her scoring guide. So she may ask students to think about the quality of their details.)
2. Before meeting with you (or submitting their work for your feedback) have students use the scoring guide to identify what aspects of quality are present in a particular piece of their work. Encourage them to use the language of the scoring guide.
3. Have them follow the same procedure to identify one or two aspects of quality they think need work.
4. Offer your feedback. If you agree, it's simple. If you can, point out a strength the student overlooked. Add to or modify what the student needs to work on, if needed.
5. Ask students to take their own and your opinions into account and decide what to do next. At first students may set large, unmanageable, or nonspecific goals. Help them, if needed, focus their plan on what is doable in the short term.
6. If your students have practiced giving formative feedback, encourage students to use each other as feedback providers.

Offering Written Feedback

You can also use the form in Figure 7.13 as a vehicle for offering written feedback. The student completes the top information and “My Opinion” and turns it in with the work. You fill out the teacher portion as you are reviewing the work, hand it back, and then the student fills in the plan. You may wish to meet with those students whose opinions differ sharply from yours.

Highlighting the Scoring Guide

This idea, from Shannon Thompson (personal communication, 2001), a curriculum specialist in the Central Kitsap School District, is another quick way to offer descriptive feedback, while also getting students to think in terms of their work's strengths and weaknesses. First, students self-assess by highlighting words and phrases on the scoring guide that describe their work. Let's say they use a yellow highlighter. You collect the work with highlighted scoring guides attached and review it. You mark your own judgment of the work's

Figure 7.13 Using Feedback to Set Goals

<p>TRAIT(S): _____ NAME: _____</p> <p>NAME OF PAPER: _____ DATE: _____</p> <p>MY OPINION My strengths are _____ _____</p> <p>What I think I need to work on is _____ _____</p> <p>MY TEACHER'S OPINION Strengths: _____ _____</p> <p>Work on: _____ _____</p> <p>MY PLAN What I will do now: _____ _____</p> <p>Next time I'll ask for feedback from: _____ _____</p>
--

Source: Adapted from *Assessment FOR Learning: An Action Guide for School Leaders* (p. 193), by S. Chappuis, R. J. Stiggins, J. Arter, and J. Chappuis, 2004, Portland, OR: Assessment Training Institute. Copyright © 2006, 2004 Educational Testing Service. Adapted by permission.

strengths and weaknesses using another color highlighter, let's say blue. Areas where the two colors merge, turning green, represent agreement. Words and phrases where the two colors remain separate represent additional feedback. You may wish to meet with those students whose scoring guides have a significant number of phrases remaining yellow and blue to clarify their understanding of quality. This strategy will work well only if students have had enough practice evaluating anonymous samples to understand the language of the scoring guide.

How Do These Activities Benefit You and Your Students?

Engaging in self-assessment prior to receiving feedback and in action planning afterwards shifts the primary responsibility for improving the work to the student, where it belongs. If you use this as part of your formative conferences with students, you will notice that gradually you will have fewer students to meet with and more students thinking about how the elements of quality you are teaching relate to their own work. In either written or verbal feedback situations, if you are going to the trouble of providing feedback, you want it to be used. Students are more likely to understand and act on your suggestions because you have asked them to access prior knowledge, which provides a mental “hook” for new information.

Strategy 4: Teach Students to Self-Assess and Set Goals

If students have had experience using rubrics following the first three strategies, when it comes time for them to self-assess, they will be prepared to do the kind of in-depth objective thinking about the quality of their work. All we as teachers have to do is provide the time and the opportunity to do so. (We discuss self-assessment and goal setting in depth in Chapter 12.)

How Can I Close the Gap?

Strategy 5: Design Lessons to Focus on One Aspect of Quality at a Time

Sometimes students must master numerous or complex elements of quality and cannot attend to all the elements at one time. So, our fifth strategy suggests that we zero in selectively in our daily teaching. One of the added benefits of creating a bulleted list to represent each trait in a rubric is that, if well crafted, the list defines relatively narrow components of quality that lend themselves to individual teaching and practice. Consider a science rubric consisting of three traits—designing the experiment, collecting and reporting data, and drawing conclusions—where the first trait includes the quality of the hypothesis. We can

teach focused lessons on how to create a high-quality hypothesis, and students can focus on just this aspect of quality until they understand it.

Strategy 6: Teach Students Focused Revision

A logical outgrowth of students learning about how to create high-quality hypotheses is that they will be able to turn to previous hypotheses they have written and revise them. Strategy 6 gives students practice at revising their own and others' work.

For example, in a writing class, after students evaluate a piece of their own work they can create a revision plan, telling how they would make their paper stronger for a particular trait. In a mathematics class, students may evaluate an anonymous work sample and work in pairs to write a letter to the (anonymous) author describing how to strengthen the solution for a given trait.

Students' revisions are based on their understanding of the scoring guide, based on our descriptive feedback, based on their self-evaluation, and based on the tips we have offered through direct instruction. They are not stuck, hands waving in the air, waiting until we can get to them, and then mumbling the mantra, "I don't get it."

Strategy 7: Engage Students in Self-Reflection, and Let Them Keep Track of and Share Their Learning

The language of the rubric becomes the language of self-reflection as well. In reflecting on how they have grown in achievement, students can use the concepts, terms, and phrases representing levels of quality to describe their journey as a reader, writer, math problem solver, social studier, and so forth. Portfolios, as we describe in Chapter 11, are an effective vehicle for collecting and sharing evidence of growth, or achievement, and of accomplishments to celebrate.

Table 7.8 presents a summary of the seven strategies and their rationales.

Table 7.8 Assessment for Learning—A Practical Performance Assessment Application

	Strategy	Rationale
1	Provide an understandable vision of the learning target. Teach students the concepts underpinning quality in your scoring guide by asking them what they already know (What makes a good ___?), then show how their prior knowledge links to your definition of quality.	Showing the connection between new information and knowledge students already have helps it all make sense and provides a link to long-term memory. It also lays the foundation for students understanding the upcoming learning.
2	Use models of strong and weak work. <ul style="list-style-type: none"> • Share anonymous strong and weak student work. Have students use the scoring guide to evaluate the samples, then share their reasons, using the language of the scoring guide. • Share published strong (and weak, if available) work. Let students comment on the quality of published examples and your own work, using the language of the scoring guide. • Share your own work. Model the “messy underside” of creating the performance or product for students. 	Student performances improve when they understand the meaning of quality. This strategy teaches students to distinguish between strong and weak products or performances, and to articulate the differences. It also encourages teachers to share aspects of the beauty of their discipline. What does it look/sound/feel like when it's done especially well? Modeling the messy underside for students reassures them that high-quality work doesn't always start out looking like high-quality work. As teachers, we tend to smooth over this part, so when the going gets messy for students, they may infer they are “doing it wrong.” What does high-quality work look like at its beginning stages? Model it.
3	Offer descriptive feedback instead of grades on practice work, pointing out what students are doing right as well as what they need to work on, using the language of the scoring guide.	Students need descriptive feedback while they're learning. It tells them how close they are to reaching the target, and it models the kind of thinking we want them to be able to do, ultimately, when self-assessing.

Table 7.8 (Continued)

4	<p>Teach students to self-assess and set goals. Ask them to identify their own strengths and areas for improvement, using the language of the scoring guide.</p>	<p>Periodic articulation about their understanding of quality and about their own strengths and weaknesses is essential to students' ability to improve. Self-assessment is a necessary part of learning, not an add-on that we do if we have time or the "right" students. Struggling students <i>are</i> the right students.</p>
5	<p>Design lessons around the elements of quality in the scoring guide. Reorganize what you already teach and find or create lessons to fill in the gaps. Focus on one aspect of quality at a time.</p>	<p>Novice learners cannot improve simultaneously all elements of quality of a complex skill or product. If your scoring guide represents a complex skill or product, students will benefit from a "mini-lesson" approach, wherein they are allowed to learn and master a portion at a time.</p>
6	<p>Teach students focused revision. Let students work in pairs to revise anonymous samples. Once they have evaluated a weak sample, ask them to use their reasons to go further: What could you do to make this receive a higher score?</p>	<p>Students need the opportunity to practice using the scoring guide as a guide to revision. That way, they, not their teachers, are doing the thinking about revision and the learning.</p>
7	<p>Engage students in self-reflection. Let them keep track of and share what they know.</p>	<p>Any activity that requires students to reflect on what they are learning and to share their progress with an audience both reinforces the learning and helps them develop insights into themselves as learners. By reflecting on their learning, students are learning more deeply and will remember it longer.</p>

TRY THIS**Activity 7.10 Create a Student-Involved Performance Assessment Plan**

Select a rubric as the basis for your plan. It can be one of your state rubrics, another one you currently use, or one you plan to use. The rubric should be a general one, that is, it should apply to more than one task—ideally to a constellation of tasks representing an important reasoning, skill, or product learning target in your content area.

Plan how you would use this rubric as a teaching tool. Begin with Strategy 1 and decide which of the recommendations would fit your subject, rubric, and student needs. Proceed through the strategies, making modifications as needed. Planning forms for this activity can be found on the CD in the file, “AFL Plan.”

If you are working as a part of a learning team, you may wish to share your plans with each other. Suggestions for how to facilitate a sharing session are also included on the CD in the file, “AFL Plan.”

DEEPEN UNDERSTANDING**Activity 7.11 Watch Video, *Designing Performance Assessments for Learning***

Judy Arter and Jan Chappuis show how to evaluate rubrics and performance tasks for quality. This video also includes a segment on how to develop a rubric, with examples taken from the work of a team of science teachers. If you are working with a learning team, watch the video, do the activities, and evaluate one or more of your own tasks and rubrics together.

Summary

Performance assessment is assessment based on observation and judgment—we observe or review a performance or product and make a judgment about its quality. The challenge with this type of essentially subjective assessment is to make it as objective as possible. Performance assessments consist of two parts: a task—what we ask the students to do—and criteria—the basis for judging quality. Much of the work in making performance assessment as objective as possible comes in the area of refining the criteria to maximize rater agreement.

Performance assessment is well suited to evaluating reasoning, skill, and product learning targets.

To select high-quality performance criteria, we look at four dimensions of quality: *Content*, *Clarity*, *Practicality*, and *Technical Quality/Fairness*. Do the criteria cover features of work that really define quality? Are each one of these features defined clearly and illustrated with models at all levels of quality? Is the rubric practical for teachers and students to use? Is there anything in the criteria that might disadvantage any group of students?

The steps in developing rubrics and performance criteria that have maximal usefulness in the classroom are (1) establish our knowledge base, (2) gather samples of student performance, (3) sort the samples by level of quality and describe the features of the work at each level, (4) cluster the features into traits, (5) identify good examples to illustrate each level, and (6) revise the rubric as it is used.

To select high-quality performance tasks, examine them for five dimensions of quality: *Content*, *Clarity*, *Feasibility*, *Fairness and Accuracy*, and *Sampling*. Does the content of the task match our learning targets and performance criteria? Is it clear to students what they are supposed to do? Can the task be carried out within the time allowed given the materials at hand? Is there anything in the task that might disadvantage any particular student or group of students? Do we have enough tasks that cover enough dimensions of the targets to ensure that we will be able to infer overall level of student mastery of the target?

To develop performance tasks, we follow the steps we used in creating extended response exercises: specify the learning to be demonstrated, specify the conditions—materials and constraints, and include the criteria by which student work will be judged. We make a series of design decisions along the way: One right answer or more than one? Written, oral and/or visual instructions, activities, and responses? Student choice or no

student choice? Individual work or group work? Spontaneously occurring event or separate event? Timed or untimed? Our answers to each of these questions are dependent on the nature of the learning target to be assessed, how we intend to use the information, and the needs of our students.

The rubrics associated with performance assessments provide the classic example of how to involve students in assessment. Rubrics can be used to help students understand where they are going, where they are now, and how to close the gap. With respect to understanding where they are going, good rubrics define quality so that students can see it. They provide a vocabulary for talking about features of quality work. Using models of anonymous strong and weak performance not only helps students deepen their understanding of the features of a quality performance or product, but also allows students to become accurate raters of performance. This accuracy is essential before students begin to self-assess.

Rubrics are also very helpful in allowing students to know where they are now and how to improve. Teachers can use them to provide descriptive feedback to students. They can also be used by students for self-assessment, goal setting, and self-reflection. This chapter presented extensive examples of each of these uses.

■ *Tracking Your Learning—Possible Portfolio Entries*

Any of the activities included in Chapter 7 can be used as portfolio entries. Remember, the learning targets for this book are outlined in Figure 7.1, 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.

DEEPEN UNDERSTANDING

Activity 7.12 Term Paper Assignment

Read the term paper assignment case study in Figure 7.14. Think about the following questions and discuss them with your team:

1. If you were a student who worked this hard to receive this feedback, how would you react?
2. What went wrong here?
3. Which problems could have been avoided? How?

TRY THIS

Activity 7.13 Help Students Understand Performance Criteria

We can use examples from everyday life to help students understand the concept of levels of quality in preparation for helping them understand the use of rubrics to self-assess. This example uses pizza. Specifically, students must answer the following questions: “What makes a great pizza?”; “What makes a good pizza?”; “What makes a minimum pizza?”; and “What makes a pizza that is below standard?”

The pizza rubric drawn by a first-grade student is shown in Figure 7.15. Think about or discuss the following questions:

1. What evidence in the rubric indicates that this student understands the concept of levels of quality?
2. What questions would you ask to help students distinguish between levels of quality of pizza if you were to ask them to build a scoring guide for pizza?
3. If you were to engage your students in a similar rubric-building activity, what food, familiar item, or process would be appropriate in your setting? The key here is to use something that students have experience with and know well so that all can contribute their ideas.
4. How might you build on this activity to take students into an academic learning target? What might be a good starting academic learning target for your students?

Figure 7.14 Term Paper Assignment Case Study

Mr. Jones, an experienced high school English teacher, gave his class the following assignment: "Read four novels by the same author and write a literary term paper arising from that experience. Develop a guiding thesis and use insights derived from your reading to defend your thesis." He had been covering American literature for decades and had been assigning term papers in this way for as long. It always worked well.

Marissa, an avid reader, had no trouble finding a socially conscious author and searching out and reading four compelling novels about the justices and injustices of our culture. The author was a woman and her stories focused on the female experience in United States history—the roles, challenges, and triumphs of women. However, Marissa had no confidence as a writer, even though her parents had told her that her writing showed talent. She wasn't buying it.

The assignment contained no information about the attributes of a good term paper. "Just apply what you've already learned," said Marissa's teacher. "Think of it as a term paper like all the others." The problem, however, was that Marissa had received almost no prior instruction in how to organize, let alone compose, such a piece. Nevertheless, she picked a prominent character from each novel and structured her paper around a comparative analysis of these women. She established the standards of comparison up front and examined key similarities and differences between and among them. To conclude, she used her comparison to speculate about the character and experience of the author.

She had to turn the draft in by a specified date or have her grade for the project reduced. She met the deadline only to discover that she wouldn't get it back—after all how could one teacher review 180 drafts! (A valid point—especially when there would be another 180 final versions to read and evaluate later.) But the teacher assured the class that as the final deadline approached, they would thank him for requiring the preliminary version.

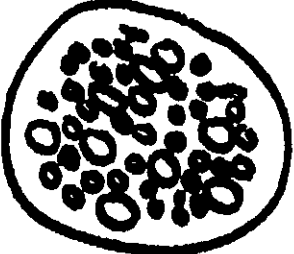
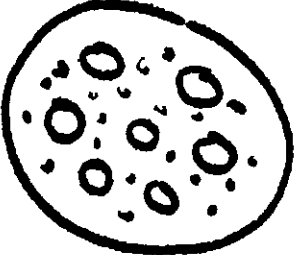
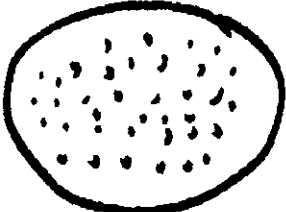


Over the next two weeks, Marissa worked to polish her paper. She revised and edited slightly—reading paragraphs to her parents and worrying that it just wasn't good enough. Finally the due date arrived and Marissa turned her paper in.

Two weeks later the paper was returned. On the cover, Mr. Jones had written two things: "B+" (certainly a very good grade by most standards, especially for a first big paper) and a single comment: "You used the word 'she' entirely too many times in this paper." There was no other feedback.

When Marissa showed the paper to her parents, they asked what she herself thought of her efforts. She dropped the paper in the waste basket, wondered aloud what the teacher really thought of her work, said she needn't have wasted so much time worrying or working, and left the room. For Marissa, this product-based performance assessment was a frustrating and unfulfilling experience.

Source: Adapted from *Practice with Student-Involved Classroom Assessment* (pp. 156–157), by J. A. Arter & K. U. Busick, 2001, Portland, OR: Assessment Training Institute. Copyright © 2006, 2001 Educational Testing Service. Adapted by permission.

Figure 7.15 First-Draft Pizza Rubric

All the toppings you like Cheese Pepperoni Sauce Crust		★
Pepperoni Cheese Sauce Crust		✓+
Cheese Sauce Crust		✓
Crust Sauce		✓-
Crust		-

Source: Rubric from Pauline Jacroux, Grade 1, Aikahi Elementary School, Kailua, Hawaii, March 2000. Reprinted from *Practice with Student-Involved Classroom Assessment* (p. 159), by J. A. Arter & K. U. Busick, 2001, Portland, OR: Assessment Training Institute. Copyright © 2006, 2001 Educational Testing Service. Reprinted by permission.