

SELECTING, REVISING, OR DEVELOPING RUBRICS

In the context of performance assessment, rubrics represent the criteria for evaluating the quality of a reasoning process, a performance, or a product. You may be familiar with a variety of rubrics, some of which you may have seen used in large-scale assessments. Although rubrics used in large-scale assessments share many features of quality with rubrics designed for classroom use, those rubrics are generally designed to yield a quick, overall summary judgment of level of achievement. As a result, they don't tend to provide the level of descriptive detail about strengths and weaknesses in individual student work, and are therefore of limited usefulness in day-to-day instruction. To function effectively in the classroom, either formatively or summatively, rubrics must provide such detail.

A good classroom-level rubric serves multiple communicative and evaluative purposes:

- Defines quality
- Makes expectations clear and explicit for students
- Describes quality to parents
- Focuses teaching
- Guides interventions
- Promotes descriptive feedback to students
- Promotes student self-assessment and goal setting
- Tracks student achievement
- Makes judgments more objective, consistent, and accurate
- Improves grading consistency

If you search for rubrics on the Internet, you'll be confronted with hundreds of choices, some of which could accomplish all of the purposes listed here and many of which could not. How do you know which is which? Whether you plan to use an existing rubric or develop one of your own, it is helpful to understand the characteristics of rubrics that make them capable of serving all intended purposes. In this section, we'll introduce the terminology commonly used to describe rubrics and then examine the three dimensions of a high-quality rubric: *Content*, *Structure*, and *Descriptors*.

Rubric Terminology

You may have heard rubrics described as *holistic* or *analytic*. Those terms refer to how many scoring scales the rubric is comprised of—one or more than one. A *holistic* rubric has only one scale—all features of quality are considered together in determining a score. An *analytic* rubric has two or more scales—features of quality have been organized into separate categories and are rated separately from one another. These are sometimes known as *multi-trait rubrics*. The reasoning rubrics we examined in Chapter 6 are examples of holistic rubrics—one for each pattern of reasoning. The Oral Presentation Rubric shown in Figure 7.7 is an example of an analytic rubric.

The structure of each rubric takes the form of *criteria*, *indicators*, *levels*, and *descriptors*. See Figure 7.7 for illustrations of each of these terms.

CRITERIA. The categories of quality in an analytic rubric are known as *criteria*. Criteria represent key, independently varying dimensions of quality. Each criterion has its own rubric. You can teach each criterion separately, students can practice and receive feedback on each criterion separately, students can self-assess on each criterion separately, and you can assign grades to each criterion separately, if desired. Criteria are also sometimes called *traits*.

INDICATORS. Criteria for complex performances or products can be broken down further into subcategories called *indicators*. Indicators are the bulleted list of features assessed in each criterion. Occasionally indicators are represented on the rubric as subheads dividing the descriptors of quality. All criteria have indicators, but not all criteria call them out with structured subheads within a rubric.

LEVELS. *Levels* on a rubric are the points on a scale defining degrees of quality. They can be labeled with numbers (e.g., 1–5), phrases (e.g., “Just Beginning”; “Halfway There”; “Success”), and/or symbols representing “basic” to “proficient” (e.g., parts of a hamburger, ice cream cone, or pizza). On an analytic rubric each criterion generally has the same number of levels.

DESCRIPTORS. *Descriptors* refer to the sentence or phrases representing each indicator at each level. Descriptors provide the details used to flesh out the indicators and differentiate the levels. In assessment *for* learning applications, the descriptors function as diagnosis and feedback about student strengths and weaknesses within the criterion.

Content of the Rubric

The *content* of the rubric defines the elements of quality essential to achieve the intended learning target. What does it assess? What are we looking for in a student’s product or performance? What will “count?” We examine this characteristic first when selecting rubrics. If the rubric under consideration falls seriously short on content, there is no need to consider it further.

A good rubric defines the intended learning target by describing what is required to do it well. If the rubric misrepresents the intended learning, students will work toward producing evidence of something other than what is desired—what students see on the rubric is how they will define quality. To make sure a rubric’s content is in good shape, we pay attention to two factors, target alignment and match to essential elements.

TARGET ALIGNMENT. Just as the task should align to the learning target(s) to be assessed, so should the rubric. The rubric’s criteria and descriptors should not focus

FIGURE 7.7 Structure of a Rubric

To illustrate the structure of a rubric, we will use an oral presentation rubric as our example.

CRITERIA

The oral presentation rubric has four *criteria*:

1. Content
2. Organization
3. Delivery
4. Language Use

INDICATORS

Each of the four criteria has several *indicators*:

1. Content
 - Clear main topic
 - All information is important to understanding the topic
 - Facts, details, anecdotes, and/or examples make topic come alive for audience
2. Organization
 - Opening introduces topic and catches audience's interest
 - Sequence of ideas supports meaning and is easy to follow
 - Transition words guide audience
 - Conclusion wraps up topic and leaves audience feeling satisfied
3. Delivery
 - Maintains eye contact with audience throughout presentation
 - Voice is loud enough for audience to hear
 - Articulates clearly
 - Speaks at a pace that keeps audience engaged without racing
 - Avoids "filler" words ("and," "uh," "um," "like," "you know")
 - Uses gestures and movement to enhance meaning
 - Uses notes only as reminders
 - Visual aids and props, if used, add to meaning
4. Language Use
 - Chooses words and phrases to create a clear understanding of the message
 - Uses language techniques (e.g., humor, imagery, simile, and metaphor) effectively as appropriate to topic, purpose, and audience
 - Explains unfamiliar terminology, if used
 - Matches level of formality in language and tone to purpose and audience
 - Uses words and phrases accurately
 - Uses correct grammar

LEVELS AND DESCRIPTORS

Each criterion has a separate scoring scale, divided into *levels*. The oral presentation rubric has three levels. Each indicator for each criterion is fleshed out into one or more *descriptors* at each level. Here is what the rubric looks like for the criterion of *Content*. Each of the other three criteria also has a rubric organized the same way.

ORAL PRESENTATION CRITERION 1: CONTENT**5: Strong**

- My presentation had a clear main topic.
- All of the information in my presentation related to and supported my topic.
- The information I included was important to understanding my topic.
- I chose facts, details, anecdotes, and/or examples to make my topic come alive for my audience.

3: Part-way There

- My topic was fairly broad, but the audience could tell where I was headed.
- Most of my details related to and supported my topic, but some might have been off-topic.
- Some of my information was important, but some details might have been too trivial to be included. Maybe I should have left some details out.
- Some of my information may not have been interesting or useful to my audience.

1: Just Beginning

- I wasn't sure what the focus of my presentation was, or I got mixed up and changed topics during my presentation. I think I wandered through a few topics.
- I didn't really know how to choose details to share, so I just used whatever came into my mind.
- I forgot to think about what information might be most interesting or useful to my audience.

Source: Adapted from *Seven Strategies of Assessment for Learning* (p. 194), by J. Chappuis, 2009, Upper Saddle River, NJ: Pearson Education. Adapted by permission.

on features that do not contribute to doing well on the learning target. Sometimes rubrics stray in their focus, as when students are asked to produce a poster to demonstrate a reasoning learning target and the rubric includes features of the poster more closely related to direction following or art than to the reasoning target. “Home-grown” rubrics, absent careful consideration of the intended learning, often suffer from this problem. If we are evaluating reasoning, the rubric should represent levels of quality for the reasoning target. Features unrelated to the learning targets should be left out or assessed separately for another purpose.

Judge the performance (demonstration) or the product (artifact) only if the performance or product is specified in the learning target. If a performance or product is not called for, make sure the rubric measures the learning target.

FAQ 7.3

Length of Rubrics

Question:

Aren't shorter rubrics better? Shouldn't the rubric fit on one page?

Answer:

Whether a rubric fits on one page depends on the use to which it will be put and the complexity of the learning target. For a rubric with more than one criterion, each criterion needs enough descriptive detail so that teachers can be consistent in judging student work and students can understand the strengths and/or weaknesses that underlie each score point.

With a lengthy rubric keep two points in mind. First, you don't have to evaluate all criteria for each piece of work. One of the strengths of multicriteria rubrics is that you can teach to, and assess, one aspect of quality at a time. Second, once students are familiar with a rubric, you can use a list of indicators for each criterion to remind them of the definition of quality, keeping the whole rubric handy for reference.

We would not advocate trading clarity for conciseness in either summative or formative applications.

Source: Adapted with permission from *Creating and Recognizing Quality Rubrics* (p. 43), by J. A. Arter and J. Chappuis, 2006, Upper Saddle River, NJ: Pearson Education. Adapted by permission.

MATCH TO ESSENTIAL ELEMENTS. The rubric's criteria and descriptors should also represent best thinking in the field about what it means to perform well on the intended learning target. Everything of importance for students at your level should be included. Three unfortunate things happen when important things are omitted: (1) we send the message that what is left out is unimportant; (2) we generate incomplete information on which to plan future instruction; and (3) we provide no feedback to students on the quality of valued elements.

By the same token, trivial features, those not important to success, *should* be left out. Some rubrics with a problem here require one way of demonstrating the learning, but the requirement may not be essential to demonstration of quality for the learning target. These rubrics inaccurately limit the definition of what it means to do well and penalize students who achieve the intended learning through a different path. If the feature is essential to the learning target, leave it in. If not, consider taking it out.

FIGURE 7.8 Characteristics of a Good Rubric**Content of the Rubric**

- *Target Alignment:* Focuses on features that contribute to doing well on the learning target.
- *Focus on Essential Elements:* Represents best thinking in the field about what it means to perform well on the intended learning target.

Structure of the Rubric

- *Number of Criteria:* Sufficient to reflect the complexity of the learning target and its intended use.
- *Independence of Criteria:* If multiple criteria, they are independent of one another.
- *Grouping of Descriptors:* If multiple criteria, descriptors are grouped logically.
- *Number of Levels:* Fits the complexity of the target and intended use of the data.

Descriptors in the Rubric

- *Kind of Detail:* Wording is descriptive of the work and can be used diagnostically in describing strengths and weaknesses.
- *Content of Performance Levels:* Levels of performance quality are parallel in content.
- *Formative Usefulness:* Language can function as effective feedback to the student and the teacher.

Structure of the Rubric

Structure refers to how the rubric is organized: criteria are defined that represent important dimensions of quality. A good rubric organizes the criteria and its associated descriptors in ways that make it possible for the user to create an accurate picture of strengths and weaknesses. Good clear structure contributes to ease of use. To maximize a rubric's structure, we pay attention to four factors: number of criteria, independence of criteria, grouping of descriptors, and number of levels.

NUMBER OF CRITERIA. The number of criteria should be sufficient to reflect the complexity of the learning target and its intended use. If the rubric is holistic, the single scale needs to sufficiently represent all important parts of the target in one scale. If the target is complex, the rubric needs to include whatever number of criteria is needed appropriately define all important categories of proficiency.

INDEPENDENCE OF CRITERIA. If there are multiple criteria, they should be independent of one another. The same or similar descriptors should appear in only one criterion. When the same feature is rated in more than one criterion, it may indicate that the criteria aren't separable. If they are separable, they can be rated independently and each feature should appear in only one criterion.

GROUPING OF DESCRIPTORS. If there are multiple criteria, all descriptors should fit under the criterion they are assigned to. In other words, the categories defined by the criteria should suit the descriptors contained within them. Grouping of descriptors is a classification challenge. If descriptors don't fit where they're placed, they should be moved, or the categories should be redefined.

NUMBER OF LEVELS. The number of levels of proficiency defined within each criterion should fit the complexity of the target and intended use of the data as well. The levels should be useful in diagnosing student strengths and next steps: there should be enough of them to reflect typical stages of student understanding, capabilities, or progress. However, there should not be so many levels that it is difficult or impossible to define each or to distinguish among them.

When a scoring guide is comprised of several categories, each of which is assigned a number of points, it is sometimes referred to as a scoring rubric. In truth, it is often a list of criteria with few or no descriptors and lots of levels. Here is an example:

Ideas and Content: 10 points

Organization: 20 points

Word Choice and Sentence Structure: 10 points

Conventions: 20 points

When a scoring guide looks like this, it's not a rubric because it's not a description of levels of quality. It is just a way to assign points, and not a very reliable one. If there are 20 points possible for Organization, there are 20 levels of quality with no guidance on how to differentiate among them—what's the difference between a 13 and a 14?

Descriptors in the Rubric

The *descriptors* are the “goes-unders”—the detail that fleshes out each level. A high-quality rubric includes descriptors that accurately represent the criteria, are complete, and are clear enough so that teachers and students are likely to interpret them the same way. To evaluate a rubric's descriptors we pay attention to three factors: kind of detail, content of the levels, and formative usefulness.

KIND OF DETAIL. The wording should be descriptive of the work. To be used formatively, a rubric's descriptors should be helpful in defining levels in ways that diagnose student strengths and weaknesses. Evaluative language should not be used in the descriptors. When a rubric uses only these kinds of evaluative terms to differentiate levels, it offers no insight into why something is good or not; it just repeats the judgment of the level.

Also, if counting the number or frequency of something is included in a descriptor, we need to ensure that changes in such counts *are* indicators of changes in quality. Under

FAQ 7.4**Number of Levels****Question:**

Shouldn't we always use an even number of levels? With an odd number, won't it be too easy to gravitate toward the middle?

Answer:

The number of levels depends on the learning target being assessed and the intended use of the information the rubric yields. Raters can easily be trained to avoid the problem of over-assigning the middle score.

Some simpler learning targets can truly only be divided into three levels of proficiency, so it makes sense to have only three levels. Others, such as the criterion *Content* in the Oral Presentation example in Figure 7.7, can be divided into five levels. This rubric has three defined levels, but it is easy to see that a student's performance might contain some descriptors from Level 5 and some descriptors for Level 3, or some descriptors from Level 1 and some from Level 3. With such rubrics, when a performance falls between two score points, highlight the phrases that describe it from each of the defined levels and then assign it the intermediate score, e.g., 4 or 2.

Source: Adapted with permission from *Creating and Recognizing Quality Rubrics* (p. 36), by J. A. Arter and J. Chappuis, 2006, Upper Saddle River, NJ: Pearson Education. Adapted by permission.

the guise of increasing the objectivity of scoring guides, it is tempting to count things—the number of sentences, the number of pieces of information, the number of topics covered, and so on. But this backfires when quality is not primarily defined by quantity. See For Example 7.3 for a sample of descriptive, evaluative, and quantitative language.

CONTENT OF LEVELS. The levels of the rubric should be parallel in their references to keys to quality. If an indicator of quality is referred to in one level, it should be mentioned in all levels. If an indicator is missing at one or more levels, there should be a logical rationale. For example, in a writing rubric, if “focus” is described at the “strong” level, it should also be described at all of the other levels. Or in a mathematics problem-solving rubric, one indicator may be “reasonableness of the solution.” In this case, even though you may have a three-level rubric, you may only describe reasonableness of the solution at the “strong” and “weak” levels, because the solution is either within a reasonable range or it isn't.

For Example 7.3

Descriptive, Evaluative, and Quantitative Language

One criterion of a rubric for a science report may include descriptors for display of information. These three examples show the differences among descriptive language, evaluative language, and quantitative language. We recommend whenever possible to use descriptive language.

Descriptive language:

- 4:** Display of information is accurate, complete, and organized so that it is easy to interpret.
- 3:** Display of information is accurate, mostly complete, and is mostly organized so that it is easy to interpret. It may have one or two small omissions.
- 2:** Display of information is partially accurate, partially complete, and may have some organization problems.
- 1:** Display of information is inaccurate, incomplete, and not well organized.

Evaluative language:

- 4:** Excellent display of information
- 3:** Good display of information
- 2:** Fair display of information
- 1:** Poor display of information

Quantitative language:

- 4:** Displays four pieces of information
- 3:** Displays three pieces of information
- 2:** Displays two pieces of information
- 1:** Displays one piece of information

Source: Adapted from *Seven Strategies of Assessment for Learning* (p. 39), by J. Chappuis, 2009, Upper Saddle River, NJ: Pearson Education. Adapted by permission.

FORMATIVE USEFULNESS. If the rubric is intended for formative use, its levels and descriptors should function as effective feedback to the student and the teacher, leading to clear conclusions about strengths and areas needing work that provide sufficient detail to guide further learning. Students should be able to use the ingredients of the rubric to self-assess, to revise their own work, and to plan their own next steps in learning. They should be able to use the criteria, levels, and descriptors to

FAQ 7.5**Including Task-specific Requirements in the Rubric****Question:**

The tasks I assign have very specific requirements and I include them in the rubric to make sure students do them and get credit for them. Can I still do this?

Answer:

It depends. If the requirements relate directly to a definition of quality regarding the learning target, they may legitimately belong in the rubric. However, if the requirements have to do with following directions—e.g., includes three characters, has five sentences—unless they are integral to the successful demonstration of the intended learning, they don't belong in the rubric and may not belong in the task.

offer one another feedback. Teachers should be able to use the rubric to determine what to teach next, to identify needs for differentiated instruction, or to identify topics for whole-group reteaching. All of the student uses require a version of the rubric in language they can understand—a student-friendly rubric.

Process for Developing Rubrics

Often existing rubrics can be revised to meet standards of quality. However, when a survey of available rubrics yields no promising candidates, you will need to start from scratch. Because you will want to use the rubric both formatively and summatively in the classroom, we recommend that you develop *general* rather than *task-specific* rubrics.

The process of rubric development involves collecting samples of existing rubrics, brainstorming features of quality, creating a draft, examining and scoring student work with the draft, and revising it. We have organized this path into six steps which we strongly recommend be carried out by teachers *working as a team* (Figure 7.9):

1. Establish a knowledge base.
2. Gather samples of student performances or products.
3. Sort student work by level of quality.
4. Cluster the descriptors into traits.
5. Identify samples that illustrate each level.
6. Test the rubric and revise as needed.

STEP 1: ESTABLISH A KNOWLEDGE BASE. To create a rubric that meets standards of quality, we have to be clear ourselves about what the performance or product looks like when it is done well. If you're an expert in creating the product or performance

FIGURE 7.9 Steps in Rubric Development

1. Establish a knowledge base.
2. Gather samples of student performances or products.
3. Sort student work by level of quality.
4. Cluster the descriptors into traits.
5. Identify samples that illustrate each level.
6. Test the rubric and revise as needed.

yourself, you may be able to work alone, but we recommend going through this process with a team as mentioned previously. A rubric development team should include some level of expertise. If you are not an expert at it, make sure you are working with someone who has experience with creating the performance or product.

As a team, begin by listing what you believe to be the characteristics of a high-quality performance or product, as called for by the learning target the rubric is intended to assess.

Next, collect as many existing rubrics as you can. These documents may provide you with inspiration and rubric language. Review the rubrics and add to your list any characteristics that you believe should be added.

STEP 2: GATHER SAMPLES OF STUDENT PERFORMANCES OR PRODUCTS. Gather a range of student performances or products that illustrate different levels of quality on the intended learning target. Good sources include your own students' work, your colleagues' students' work, books on teaching your subject, your State Department of Education's website, and other Internet sites. If the learning target requires a performance, this will require gathering audio or videotaped examples.

In general, try to gather at least 20 samples representing more than one topic or task. Using samples from only one topic or task may lead you to develop a rubric that is too task-specific to be useful across tasks. A variety of samples helps ensure that all important criteria are included on the final rubric.

A note of caution: If the samples come from your own school, once you are finished with this process, don't publish them as "anchors" for the rubric. Any published anchors require permission from the student and more anonymity than can be guaranteed in one school.

STEP 3: SORT STUDENT WORK BY LEVEL OF QUALITY. Begin by examining the samples of student work and sorting them into three stacks representing your evaluation of them as Strong, Medium, or Weak. Write down your reasons for placing each sample in each stack as you go. Have each member of the team do this independently.

The goal of sorting is not to get every sample in exactly the correct stack. The goal is to develop a complete list of the *reasons why* a sample should be placed in a particular stack. Eventually, the rubric you develop will be capable of judging samples accurately and consistently, but for now, focus on being descriptive in your reasons.

For each sample, write down exactly what you are saying to yourself as you place it into the stack. Don't wait until you have the samples sorted—it's harder later to remember all of the details of your reasons. Include as much detail as you can. Dig below general statements to a description of the evidence that leads to the general conclusion. For example, "logical" might be on your list of desired characteristics, but that is a general statement and does not describe the features present that lead you to determine whether or not a solution or argument is logical.

As another example, if "fluency" is on your list, 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. The descriptive phrases that you create now will form the core of your rubric descriptors.

Then, as a team, compile all of your descriptions of the samples in each stack. 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 each level. We recommend trying to come up with as broad and long a list as possible.

As people sort samples, they sometimes discover that three levels is too few. Their eye develops to the point that they find work that is between two of the levels and they want to have four to six stacks. This is the beginning of determining your final number of levels. If three levels seem adequate, that's your number. If you can distinguish more than three independent levels, sort student work into the number of stacks you think you need and create the list of descriptors for each level. That is fine as long as you can find descriptors and/or sample performances that differentiate the levels. As long as you and students can differentiate performance levels, identify however many performance levels you need.

STEP 4: CLUSTER THE DESCRIPTORS INTO TRAITS. Your sorting and describing will result in a hodgepodge of descriptors at each level of performance. Some descriptors will be closely linked and can be assigned to a category: someone will say "Wait a minute. We have a whole lot of statements that refer to fluency. Why not group them together?" Some descriptors will overlap: you may hear, "Wait a minute. 'Speaks in paragraphs' is the same as 'Combines several sentences.' Why not delete one?"

FAQ 7.6**Student-developed Rubrics****Question:**

To maximize student motivation, shouldn't we develop our rubrics with them?

Answer:

Involving students in developing criteria has several advantages and some drawbacks. It helps them internalize the standards of quality and it helps us refine our thinking. But most often we as teachers know more about the criteria for quality than students do, so developing criteria with students requires a robust sample set to help them refine their vision. For example, 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.

A teacher should not begin instruction aimed at helping students perform well on a performance assessment without a clear vision of what the rubric should contain. The role of instruction is to bring students to a sufficient understanding of the keys to quality to be able to perform well. If you want to develop criteria with students, be prepared to assist them in their discovery of good criteria through use of thoughtfully chosen examples. Leading students through a process of devising a student-friendly version of the already-created rubric is one excellent instructional strategy. But the original articulation of keys to quality is the responsibility of the teacher.

Other descriptors may need to be separated into two categories: "I had trouble placing a student performance in a single category because it was strong in fluency but weak in pronunciation. Let's score those two dimensions separately."

This is the beginning of an analytical structure for the rubric.—when you can sort out broad, independently varying categories of strengths and weaknesses, these indicate separate criteria. Once you have drafted the categories, it's time to refine them. You might decide that two criteria really refer to the same thing, or that one criterion should be divided into two or more criteria because of their independence from one another and the importance of being able to assess each separately. Most rubrics go through several stages of criteria definition and organization.

STEP 5: IDENTIFY SAMPLES 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

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. 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.

Here are rules of thumb for selecting samples to illustrate criteria and levels:

1. *Start with the extremes.* Identify what you consider to be classic examples of strong and weak performances or products—ones that match a good number of the descriptors in the highest and lowest categories. Choose samples that everyone on the development team can agree on. When students are first learning to analyze samples for quality, they need examples that are fairly straightforward. Leave the more ambiguous examples for later, when students have more fully developed their sense of quality.
2. *Find examples for the middle level(s).* If you are using an odd number of levels, find samples that represent the approximate midpoint in the progression to excellence for each criterion. These samples will demonstrate the partial understanding or developing proficiency that is described by the phrases attached to your Middle level. If you are using an even number of levels, you will need to find two or four sets of midrange examples. For example, if your rubric has four levels, you will select samples for each criterion that typify the descriptors at level 2 (weaker, rather than stronger) and samples for each criterion that typify the descriptors at level 3 (stronger, rather than weaker).
3. *Find several different examples across assignments that illustrate each level.* The purpose of the samples is to help with training raters to apply the criteria consistently, whether those raters are teachers or students. Teachers and students need to be able to apply the rubric across tasks, so the samples should help in learning how to do that.
4. *Keep your eye out for examples illustrating typical problems.* Select examples that illustrate common errors students make, misconceptions they are likely to have, and flaws in reasoning. Carefully selected examples of typical problems can function well as teaching tools if students will be using them to practice scoring with the rubric.

The process of finding examples of performances or products at each level for each criterion usually results in revisions to the descriptors and criteria.

STEP 6: TEST THE RUBRIC AND REVISE AS NEEDED. Now is the time to test the rubric and note how you might improve it. Score student samples with your draft rubric and ask students to score anonymous samples as well. Unless you're spectacularly

good at rubric development or spectacularly lucky, you'll identify some combination of the following problems:

1. Some student performances or products include features not mentioned in the rubric. If the features are indeed salient to quality, add descriptors and perhaps, indicators. Especially try to add descriptors that clarify general features or concepts. Don't despair that your rubric is becoming unwieldy. It's part of the revision process: we expand text to include all possible options before paring it down to the most-needed elements.
2. Features of student work seem to be rated in more than one criterion. Note this when it occurs. It might be that some descriptors are repeated across criteria. You will need to decide in which criterion the feature fits best. In some cases you may need to merge two criteria into one because they overlap to the extent that they can't be scored separately.
3. Criteria seem too broad. Sometimes there are a number of indicators in a criterion that could be separated out to form two or more separate criteria. Do this if you would like to teach, diagnose, and assess the parts separately. Make sure that the new criteria structure does represent independently varying features of quality.
4. The content of some levels is not parallel. You find that some descriptors at one level are not represented at other levels. In this case, write the descriptors for the other levels.

Evaluating the Rubric for Quality

We have developed the Rubric for Rubrics to help you evaluate any performance rubric for the degree to which it meets the standards of quality on the three criteria described in this section: *Content*, *Structure*, and *Descriptors*. The Rubric for Rubrics, Figure 7.10, can also be found in the Chapter 7 CD file.

USE STAGE

At this stage we conduct and score the assessment and revise it as needed for future use. As we noted previously, problems can still crop up even with the best planning. It is a good idea to keep notes of any potential sources of mismeasurement that may have compromised students' ability to show what they know and can do. If something appears to have gone awry and you can't identify the problem, use the Rubric for Tasks and the Rubric for Rubrics to troubleshoot the performance assessment.

FIGURE 7.10 Rubric for Rubrics

		Rubric for Rubrics: Content	
Indicator	Level 3: Ready to Use	Level 2: Needs Some Revision	Level 1: Don't Use
Target Alignment	Criteria and descriptors align directly with the content standards or learning targets they are intended to assess.	The rubric includes one or two small features that are not related to the intended content standards or learning targets.	The rubric focuses on features that are not related to the intended content standards or learning targets. One or more of the following applies: <ul style="list-style-type: none"> • The criteria and descriptors inappropriately focus on dimensions of the task rather than the learning targets. • The learning targets are not clear.
Focus on Essential Elements	Criteria and descriptors represent best thinking in the field about what it means to perform well on the content standards or learning targets. <ul style="list-style-type: none"> • Everything of importance (for students at your level) has been included. Trivial and unrelated features are left out. • If the rubric is a developmental continuum, the content represents the best thinking in the field about how proficiency develops over time. 	A few descriptors might be irrelevant or unimportant for defining proficiency, but most are relevant.	You can think of many important indicators of quality that are missing, the rubric focuses on irrelevant features, or you find yourself asking, "Why should students have to do it <i>this way</i> ?"

(continued)

FIGURE 7.10 Rubric for Rubrics (continued)

Indicator	Rubric for Rubrics: <i>Structure</i>		
	Level 3: Ready to Use	Level 2: Needs Some Revision	Level 1: Needs Major Revision
Number of Criteria	<p>The number of criteria reflects the complexity of the learning target and its intended use.</p> <ul style="list-style-type: none"> • If the rubric is holistic, a single scale sufficiently represents the target or the use is solely summative. • If the target is complex and the use is formative, the rubric is analytic; the number of criteria appropriately define categories of proficiency. 	<p>If a rubric is analytic, the number of criteria needs to be adjusted: either a single criterion should be made into two or more criteria, or two or more criteria should be combined.</p>	<p>The rubric is holistic (a single scale) when it needs to be analytic (multiple criteria) to better reflect the level of complexity of the learning targets to be assessed and the intended use of the data.</p>
Independence of Criteria	<p>If there are multiple criteria, they are independent of one another—the same or similar features are represented in only one criterion.</p>	<p>The criteria are mostly independent of one another, but in some cases features are represented in more than one criterion.</p>	<p>The criteria are not independent of one another. The same or similar features are represented in multiple criteria throughout the rubric, to the extent that the criteria do not function as separate categories.</p>
Grouping of Descriptors	<p>If there are multiple criteria, indicators and descriptors are grouped logically within each criterion. All descriptors fit under the criterion in which they are placed.</p>	<p>Most indicators and descriptors under a criterion are placed correctly, but a few need to be moved to a different criterion.</p>	<p>Indicators and descriptors that go together don't seem to be placed together; descriptors that are different are placed together; the categories don't work.</p>

Indicator	Level 3: Ready to Use	Level 2: Needs Some Revision	Level 1: Needs Major Revision
Number of Levels	<p>The number of levels fits complexity of the target and intended use of the data. There are enough levels to reflect typical stages of student understanding, capabilities, or progress, but not so many that it is impossible to distinguish among them.</p>	<p>There are a few too many levels of quality to distinguish among, so some will need to be merged; or, there are not quite enough levels to reflect typical stages of student growth, so more will have to be created.</p>	<p>The number of levels is inappropriate for the learning target being assessed or the intended use of the rubric. One or more of the following is true:</p> <ul style="list-style-type: none"> • There are so many levels it is impossible to reliably distinguish among them. • There are far too few levels to be useful in tracking student growth. • It is impossible to define the number of levels indicated.
Rubric for Rubrics: Descriptors			
Indicator	Level 3: Ready to Use	Level 2: Needs Some Revision	Level 1: Needs Major Revision
Kind of Detail	<p>Wording is descriptive of the work. There is enough detail that it is possible to match a student performance or product to the appropriate level. Descriptors provide an accurate explanation of the characteristics of quality. If counting the number or frequency of something is included as an indicator, changes in such counts <i>are</i> indicators of changes in quality.</p>	<p>Wording is mostly descriptive, but has one or more problems:</p> <ul style="list-style-type: none"> • The rubric includes a few terms that are vague. • Some language is evaluative rather than descriptive of the work. • Only the top level of quality is described sufficiently; the other levels include insufficient or no descriptive detail. • The rubric mostly avoids frequency counts, but there are a few times that counts are used even though changes in such counts do not equate to changes in level of quality. 	<p>Wording is not clear. One or more problems exist:</p> <ul style="list-style-type: none"> • Descriptors consist of vague terms without clarification, e.g., "extremely," "thoroughly," "completely," or "insightful." • Descriptors rely heavily on evaluative language to differentiate levels of quality, e.g., "mediocre," "clever," or "above average." • The rubric is little more than a list of categories and a rating scale. • Descriptors consist almost solely of counting the number or frequency of something, when quantity does not equate to quality.

(continued)

FIGURE 7.10 Rubric for Rubrics (*continued*)

Indicator	Level 3: Ready to Use	Level 2: Needs Some Revision	Level 1: Needs Major Revision
Content of Levels	The features described across the levels are parallel in content. If a feature is present in one level, it is present in all levels. If a feature is missing at one or more levels, there is a logical rationale.	The levels of the rubric are mostly parallel in content. One or a few descriptors of quality at one level are missing at other levels when they should be present.	The levels of the rubric are not parallel in content. Most descriptors of quality are not represented at all levels and there is not an adequate rationale for their omission.
Formative Usefulness	<p>If the rubric is intended for formative use, its language can function as effective feedback to the student and the teacher, describing strengths and areas needing work in sufficient detail to guide further learning.</p> <ul style="list-style-type: none"> • Students can easily use the rubric to revise their own work and plan their own next steps. • Teachers can easily translate results into instruction. 	<p>If the rubric is intended for formative use, some parts can function as effective feedback to the student and the teacher, describing strengths and areas needing work in sufficient detail to guide planning for further learning. Descriptors in other parts need refining to accomplish this purpose.</p>	<p>If the rubric is intended for formative use, it cannot be used provide effective feedback to students and teachers because it doesn't describe strengths and areas needing work in sufficient detail to guide next steps in learning.</p>

SEVEN STRATEGIES FOR USING RUBRICS AS INSTRUCTIONAL TOOLS IN THE CLASSROOM

You're all set with a great task and rubric. 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 rubric as a teaching tool to help students become competent, confident self-assessors and improve their performance in any subject.

Where Am I Going?

STRATEGY 1: PROVIDE STUDENTS WITH A CLEAR AND UNDERSTANDABLE VISION OF THE LEARNING TARGET. Motivation and achievement both increase when instruction is guided by clearly defined learning targets. Activities that help students answer the questions, "What's the learning? What am I responsible for accomplishing?" set the stage for all further formative assessment actions.

In the context of performance assessment:

Teach students the concepts underpinning quality in your scoring rubric by asking them what they already know ("What makes a good ____?"), then show how their prior knowledge links to your definition of quality.

Rationale:

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 upcoming learning.

STRATEGY 2: USE EXAMPLES AND MODELS OF STRONG AND WEAK WORK. Carefully chosen examples of the range of quality can create and refine students' understanding of the learning goal by helping students answer the questions, "What defines quality work? What are some problems to avoid?"

In the context of performance assessment:

- Use models of strong and weak work.
- Share anonymous strong and weak student work. Have students use the scoring rubric to evaluate the samples, then share their reasons, using the language of the scoring rubric.
- 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 rubric.
- Share your own work. Model the "messy underside" of creating the performance or product for students.

Rationale:

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 different 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 that they are "doing it wrong." What does high-quality work look like at its beginning stages? Model it.

Where Am I Now?

STRATEGY 3: OFFER REGULAR DESCRIPTIVE FEEDBACK. Effective feedback shows students where they are on their path to attaining the intended learning. It answers for students the questions, "What are my strengths?" "What do I need to work on?" "Where did I go wrong and what can I do about it?"

In the context of performance assessment:

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 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.

Rationale:

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.

STRATEGY 4: TEACH STUDENTS TO SELF-ASSESS AND SET GOALS. Strategy 4 teaches students to identify their strengths and weaknesses and to set goals for further learning. It helps them answer the questions, "What am I good at?"; "What do I need to work on?"; and "What should I do next?"

In the context of performance assessment:

Strategy 4 includes anything students do to identify where they are with respect to mastery of the desired learning and to set goals for improvement. Black and Wiliam (1998a) assert that for assessment to bring about learning gains, it has to include student self-assessment: "self-assessment by the students is not an interesting option or luxury; it has to be seen as essential" (pp. 54–55). In performance

assessments, you can ask students to use the rubric to identify their own strengths and areas for improvement. If you have given them descriptive feedback based on the rubric, you have modeled for them the kind of thinking they are to do when self-assessing. You can teach them to set specific, achievement-related goals and to make a plan to accomplish them.

Rationale:

Periodic articulation about their understanding of quality and about their own strengths and weaknesses is essential to students' ability to improve.

How Can I Close the Gap?

STRATEGY 5: DESIGN LESSONS TO FOCUS ON ONE LEARNING TARGET OR ASPECT OF QUALITY AT A TIME. When assessment information identifies a need, we can adjust instruction to target that need. In this strategy, we scaffold learning by narrowing the focus of a lesson to help students master a specific learning goal or to address specific misconceptions or problems.

In the context of performance assessment:

Students who are not yet proficient at creating a complex performance or product have a difficult time improving simultaneously on all elements of quality. This strategy suggests that you may want to teach lessons that address your rubric one criterion at a time. In some instances, you may need to focus on only one *part* of one criterion at a time. For example, a writing rubric might have a criterion called *Organization*. Within that criterion are descriptors about the quality of the introduction, the sequencing of ideas, transitions between ideas, pacing, and the conclusion. If your students are not writing effective and inviting introductions to their papers, give them some practice with that single aspect of the whole rubric. You could use strategies 1 through 3: Ask "What makes a good introduction?"; share examples of strong and weak introductions; have students write an introduction to something they are working on; and offer descriptive feedback based on strengths and weaknesses of introductions as described in your rubric.

Rationale:

Novice learners cannot improve simultaneously all elements of quality of a complex skill or product. If your scoring rubric 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.

STRATEGY 6: TEACH STUDENTS FOCUSED REVISION. When a concept, skill, or competence proves difficult for students, we can let them practice it in smaller segments and give feedback on just the aspects they are practicing.

In the context of performance assessment:

Any activity that allows students to revise their initial work with a focus on a manageable number of aspects of quality, problems, or learning targets is a logical next step after teaching focused lessons. Alternatively, let them create a revision plan, detailing the steps they would take to improve their product or performance, and let that stand in for the actual revision. This is especially useful in an assessment *for* learning context; students can think about revision more frequently, because each instance takes less time. Strategy 6 gives students practice using the rubric to self-assess and to guide their revisions.

Rationale:

Students need the opportunity to practice using the scoring guide as a guide to revision. When they do this, it is the students and not you who are doing the thinking about and the work of revision; this translates into deeper learning.

STRATEGY 7: ENGAGE STUDENTS IN SELF-REFLECTION AND LET THEM KEEP TRACK OF AND SHARE THEIR LEARNING. Long-term retention and motivation increase when students track, reflect on, and communicate about their learning. In this strategy, students look back on their journey, reflecting on their learning and sharing their achievement with others.

In the context of performance assessment:

Set up a system, such as a portfolio, that lets students track their learning along the way. Along with the artifacts, include the scoring rubrics you and they have used for feedback, self-assessment, and goal setting. Ask them to reflect on the contents of their portfolio, summarize their progress, and to comment on it: What changes have they noticed? What is easy that used to be hard? Where have they been surprised? Disappointed? Excited? What insights into themselves as learners have they discovered?

Rationale:

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.

(See Chapter 9 for information about how students can track their learning, Chapter 11 for how to set up and use portfolios for tracking and reflecting on progress, and Chapter 12 for the kinds of conferences students can participate in to share their learning.)

For a more detailed treatment of how each of these strategies can be used with performance assessment across grade levels and subject areas, refer to Chappuis (2009).

My Classroom Then and Now 7.1

Bruce Herzog

I used to ...

I have been an elementary school teacher for over twenty-five years. For the first fifteen years of my career I ended every year feeling disheartened by the fact that most of the students who were not meeting grade-level standards when they entered my room the previous fall were still below grade-level standards when the year ended. So each new year I worked even harder to improve my instruction because, throughout my career, I had been told over and over that good instruction would lead to good learning. I embraced new strategies and new programs of instruction that promised to lead to high achievement for all students, yet success still remained out of reach for far too many of my students.

Now I ...

When I first heard about formative assessment I was skeptical, but when I read the research that showed that effective assessment practices could lead to unprecedented achievement gains it convinced me to give it a try. I began by having my students set goals and reflect on their learning. It became immediately apparent that most students weren't exactly sure about what they were supposed to be learning. This led me to focus on making sure that the learning targets I expected students to meet were clear. I did this through the use of rubrics, study guides, and continued student goal setting and reflection. Having clear learning targets helped me focus my instruction on those clearly identified learning targets and enabled me to refine my assessments to reflect exactly what students were learning.

What I notice as a result ...

I saw almost immediate results. Within the first month that I began using formative assessment practices in my classroom I saw achievement gains from virtually all students with the greatest gains being made by the lowest-achieving students. Using these practices has completely changed the learning environment in my classroom. Good instruction is still important, but the focus has shifted from teaching to learning and from the teacher to the student. Now that learning targets are clear to students they are able to take responsibility for meeting those learning targets. By regularly reflecting on where they are in relation to the targets they are able to decide what action they need to take or what help they need in moving toward meeting those targets. Formative assessment has removed the barriers to learning in my classroom.

Do all students reach grade level standards in my room now? No, but many more do now than before I began using formative assessment practices and I am seeing dramatic improvement in those students who still have a way to go. More