Chapter 4
Establishing Criteria and Standards for Grading

MUCH OF WHAT WE DO and say in our classrooms—face-to-face or online—plays a role in letting the students get the message of what we expect of them. However, this chapter focuses on the written set of criteria and standards that a teacher presents to the students before they begin the assignment or exam and that the teacher, a department, or a general education program may use to assess student learning (departmental and general education assessment are discussed in Chapters Eleven and Twelve). We illustrate three common forms for such documents:

- Grade descriptions—what constitutes an A grade, a B, and so on
- Check sheet—lists of what the student (or peer responders) should check for and what the teacher will look for
- Rubric—a format in which each trait of the student’s work (such as evidence, organization, use of graphs) is described using a scale from high to low

The forms may blend or overlap. Each has its advantages and disadvantages.

Descriptions of Grades

One way to clarify criteria and standards for students is to describe the requirements for each grade. An example is Dorothy Solé’s description of the work required to earn points in student journals for introductory Spanish at Raymond Walters College of the University of Cincinnati (Exhibit 4.1).

Solé’s scale directly addresses students’ question, What do I need to do to get each level of points? It makes clear that merely producing correct language is not enough for full points; for the highest level, Solé values risk taking and stretching to the edge of one’s capability. Such a description can prevent students from coming to Solé to complain that they got only three points even though they made no errors.
### EXHIBIT 4.1

**Description of Point Levels for Journals in Introductory Spanish**

4. The content of the journal is comprehensible. Although there are errors, verb tenses, sentence structure, and vocabulary are correctly used. The author has taken some chances, employing sentence structures or expressing thoughts that are on the edge of what we have been studying. The entries are varied in subject and form.

3. There is some use of appropriate verb tenses and correct Spanish structure and vocabulary, but incorrect usage or vocabulary interferes with the reader’s comprehension.

2. The reader finds many of the entries difficult to understand or many entries are simplistic or repetitious.

1. The majority of the entries are incomprehensible.

---

A grade-based scale like this lumps together various factors that could explain why a student got a certain number of points. If a student got three points instead of four, did the problems lie with verb tenses, Spanish structure, usage, vocabulary, or risk taking? An aggregated grade-based scale like this must rely on the teacher’s comments to identify specific successes and problems.

Another example is a grade description that John Breihan, the history faculty member we have been following, distributes to students before they begin their argumentative essays about, for example, whether Louis XIV was a “good” king (Exhibit 4.2).

Breihan constructed the scale by examining actual student papers, and his eleven statements describe the kinds of papers he typically receives. He indicates approximate grade equivalents along the left-hand side of his sheet. When he responds to a student’s draft or finished essay, he circles one of the eleven statements that most nearly describes that students’ paper. He also makes marginal comments designed to illustrate and expand on the information offered by the statement.

Sole’s and Breihan’s grading descriptions are assignment specific. It is also possible to construct more broadly stated descriptions that can apply to more than one assignment, provided the assignments have common elements. For example, Facione and Facione (2004) offer a scoring sheet for critical thinking that describes each of four levels of critical thinking. For level 4 (the top score), the student’s performance “consistently does
**EXHIBIT 4.2**

**Breihan’s Grading Scale for Argumentative Essays in Western Civilization Course**

This scale describes the common types of paper but may not exactly describe yours; my mark on the scale denotes roughly where your essay falls. More precise information can be derived from comments and conferences with the instructor.

**Grade**

1. The paper is dishonest.
2. The paper completely ignores the questions set.
3. The paper is incomprehensible due to errors in language or usage.
4. The paper contains very serious factual errors.
5. The paper simply lists, narrates, or describes historical data and includes several factual errors.
6. The paper correctly lists, narrates, or describes historical data but makes little or no attempt to frame an argument or thesis.
7. The paper states an argument or thesis, but the argument or thesis does not address the question set.
8. The paper states an argument or thesis, but supporting subtheses and factual evidence are:
   - Missing
   - Incorrect or anachronistic
   - Irrelevant
   - Not sufficiently specific
   - All or partly obscured by errors in language or usage
9. The paper states an argument on the appropriate topic, clearly supported by relevant subtheses and specific factual evidence, but counterarguments and counterexamples are not mentioned or answered.
10. The paper contains an argument, relevant subtheses, and specific evidence; counterarguments and counterexamples are mentioned but not adequately answered:
   - Factual evidence is incorrect, missing, or not specific.
   - Linking subtheses are either unclear or missing.
   - Counterarguments and counterexamples not clearly stated; employs “straw man” argument.
11. The paper adequately states and defends an argument and answers all counterarguments and counterexamples suggested by:
   - Lectures
   - Reading assignments (specific arguments and authors are mentioned by name)
   - Common sense

all or most of the following.” The rubric then lists a number of factors that various assignments might require, including students’ management of evidence, argument, alternative points of view, conclusions, results, and procedures. Next, the authors describe a level 3 performance, including descriptions of the same factors, but at a lower level of competence.
This choice of factors allows the rubric to be applied to a variety of assignments. This broader use may be important as a department or a general education program attempts to score student work from multiple classes (see Chapters Eleven and Twelve).

Checklists

Another way to convey criteria and standards is by constructing a list of items that the teacher will use in grading the paper and that the student or peers can self-check beforehand. Checklists may simply name categories ("I will base the grade on the quality of your argument, your ability to engage your audience . . .") or describe the highest type of work the teacher expects ("Does your essay contain a clear thesis? Does it support the thesis with evidence from the texts?"). Checklists ask the student or peers to attend to items without requiring a scale or graded description. Exhibit 4.3 is a checklist that Walvoord uses for students who are submitting analyses of editorials for a course in political writing.

---

EXHIBIT 4.3

Checklist for Analysis of Newspaper Editorials

- I have chosen an editorial that takes a position on a debatable issue of public policy at the local, state, or national level.
- I have attached a copy of the editorial to this paper.
- I have summarized the editorial's main point in a few sentences. The summary is less than 10 percent of the length of my analysis.
- I have NOT focused on whether I agree or disagree with the author's position; instead, I have analyzed the editorial.
- I have analyzed the editorial in the ways we have been learning in class, including:
  - how the writer explains the background and sets the stage
  - what audience the writer appears to be addressing
  - how the writer states what she or he wants the audience to do or believe
  - how the writer tries to convince or change the audience
  - the type and quality of evidence the writer uses
  - the accuracy and integrity of any statistics or data the writer uses
  - unstated assumptions behind the writer's position
  - writer's bias or use of loaded terms
- I have evaluated the strengths and weaknesses of the editorial in terms of its effectiveness for its audience.
- I have included an evaluation of the integrity of the writer in avoiding bias, prejudice, distortion of facts, or unsupported claims.
Walvoord asks the students to put a written check mark next to each item to show they have done it to the best of their ability. Students submit the checklist with their assignment. Walvoord does not accept the assignment unless the checklist is attached. Can a student check something she has not done? Can she believe she has done something that she in fact has not done? Sure. But the checklist brings the criteria and standards strongly into mind, and it reinforces the importance of students’ self-assessing their work. Students do report that they have revised their work again after using the checklist. Such a list could also be used for peer checking.

**Rubrics**

The term *rubric* is widely used to refer to a format in which the traits of the student’s work are separately named, and each trait is evaluated according to a scale from high to low. You can find many rubrics online with search terms such as *rubric* and the name of the discipline. (See Stevens and Levi, 2005.)

Rubrics vary in their level of detail. The most minimal rubric simply names a trait with a noun or noun phrase and lists the numbers or letters of the scale. Exhibit 4.4 is an example of a minimal rubric for senior history seminar papers.

A minimal rubric is probably better than nothing. It helps faculty to articulate, and students to understand, the qualities faculty are looking for in a senior seminar paper. However, because the individual numbers are not described, a minimal rubric provides little guidance. If an individual teacher uses such a rubric, many students are still likely to ask why they got a 4 in organization. If multiple faculty use the rubric for assessing the quality of students’ senior papers as a whole, the department is likely to discover that a single paper will be rated differently by different faculty.

**EXHIBIT 4.4**

**Minimal Rubric for Senior History Seminar Papers**

<table>
<thead>
<tr>
<th></th>
<th>5</th>
<th>4</th>
<th>3</th>
<th>2</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Title</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Thesis</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Argumentation</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Historical significance of project</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Organization</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Effective Grading

EXHIBIT 4.5
Rubric for a Business Management Case Analysis: Analysis of the Firm’s Goals

<table>
<thead>
<tr>
<th></th>
<th>5</th>
<th>4</th>
<th>3</th>
<th>2</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>The statement about goals is</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>consistent with the materials in</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>the case.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>The writer presents sufficient and</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>clearly organized evidence for the</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>summary of the firm’s goals.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>The writer has chosen the most</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>important or primary goals.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

members. A minimal rubric works as a starting place, but most faculty want to move toward greater specificity.

Somewhat greater detail is provided by a rubric that describes the highest performance and then lists the scoring levels. Exhibit 4.5 shows a rubric used for students’ case analyses in a senior-level business course (the complete rubric is example 12 in Appendix A). This rubric provides more guidance than a minimal rubric because it describes what students should aim for. Nevertheless, it still leaves some questions—for example, what is the difference between a 2 and a 3 for “sufficient and clearly organized evidence”?

The most complete type of rubric describes the qualities of student work at each level of the scale. Anderson’s rubric for students’ original scientific experiment names each trait she wants to measure: for example, methods and materials, experimental design, and interpretation of results. Then she describes each level of performance. (The partial rubric is in Exhibit 4.6. The assignment sheet is in Exhibit 3.1. The full rubric is example 1 in Appendix A.)

Fully articulated rubrics such as Anderson’s require thought and hard work to construct, but they represent the most complete description of criteria and standards. They offer the best chance at interrater reliability—that is, the probability that several faculty raters will score a particular student’s work in the same way. Full rubrics force the faculty member to clarify in great detail what she is looking for. Thus, full rubrics provide explicit guidance for teaching and for students’ peer review and self-checking. (The term primary trait analysis is sometimes used to refer to full rubrics.)
EXHIBIT 4.6
Rubric with Full Descriptions of Each Scoring Level

Assignment: Design and conduct an original scientific experiment and write a report using scientific format.

Trait: Title
Level 5 Is appropriate in tone and structure to a science journal; contains all necessary descriptors for placement in a scientific database; contains necessary brand names; identifies functions of experimentation; allows reader to anticipate design.
Level 4 Is appropriate in tone and structure to science journal; contains most descriptors; identifies function and brand names; suggests aspects of design.
Level 3 Identifies function and brand name but does not allow reader to anticipate design.
Level 2 Identifies function or brand name, but not both; lacks design information or is misleading.
Level 1 Is patterned after another discipline or missing.

Trait: Methods and Materials Section
Level 5 Contains appropriate, quantifiable, concisely organized information that allows the experiment to be replicated. All information in the report can be related back to this section. Identifies sources of data. Sequences information appropriately. No wordiness.
Level 4 As above, but contains unnecessary information or wordiness.
Level 3 Experiment could be replicated from the information given. All information in the report can be related back to this section. However, fails to identify some data sources or has problematic sequencing.
Level 2 Marginally replicable. Parts of basic design must be inferred. Procedures not quantitatively described. Some information in Results or Conclusions sections cannot be anticipated by reading this section.
Level 1 Describes experiment so poorly it cannot be replicated.

See Lloyd-Jones, 1977.) The rest of this chapter discusses how to construct and use full rubrics.

How to Construct a Rubric
Exhibit 4.7 summarizes the steps for constructing a scoring rubric. If possible, work from examples of past student performances, grading checklists,
EXHIBIT 4.7
Steps for Constructing a Rubric

1. Choose a test or assignment that tests what you want to evaluate. Make clear your objectives for the assignment—what you want students to learn and do.
2. Identify the criteria, or "traits," that will count in the evaluation. These are nouns or noun phrases, such as "thesis," "eye contact with client," "use of color," or "control of variables."
3. For each trait, construct a two- to five-point scale. These are descriptive statements. For example, the trait may be "thesis." The scale for "thesis" will have several levels, each with a description of the performance for that level. For example, a thesis that scores a 5 does X, Y, and Z. A thesis that scores a 4 does X and Y. And so on.
4. Try out the scale with a sample of student work or review with colleagues and revise.

descriptions of criteria, comments on assignments or tests—and anything else that has helped you in the past to articulate criteria for student performance. The last part of this chapter addresses common questions that arise in constructing rubrics.

Using the Rubric for Scoring
Once you have a rubric, the next topic is how to score student work with it. For a quick and easy example, use Anderson’s “title” trait, set out in Exhibit 4.6, to rate each of following titles, written by Anderson’s students before she developed her rubric:

U A Comparison of Prell and Suave Shampoo
V The Battle of the Suds: Budweiser and Weiderman Beer
W Would You Eat Machine-Made or Homemade Cookies?
X A Comparison of Arizona and Snapple Ice Tea for pH, Residue, Light Absorbency, and Taste
Y Research to Determine the Better Paper Towel
Z A Comparison of Amway Laundry Detergent and Tide Laundry Detergent for Characteristics of Stain Removal, Fading, Freshness, and Cloth Strength

All titles contain fewer than twenty-five words, as the assignment sheet stated. When you try your hand at scoring these, use whole scores, not halves. To see how you did, look at Exhibit 4.8, which contains the scores Anderson gave.
After trying this scoring, you might recommend changes to the scale. Perhaps, after trying to score title Z, you would recommend that the words “is concise” be added to level 5 and “meets all criteria for 5 but may be wordy” be added to level 4. After trying to score title W, you may suggest that the scale specify whether modeling after another discipline forces a score of 1 regardless of whether the writer identifies brand names or includes other features.

Rubrics tend to be revised as you use them, and they should be. The benefit of constructing a scoring rubric lies as much in the hard thinking it requires, and in the influences it exerts on teaching and learning, as in the final score that emerges.

Why Take the Time to Do a Rubric?
How much time does it take to do a rubric? That depends on how many traits, whether you have previous grading criteria to work from, and whether you’re wired to think in this way. Faculty members working from previous grading checklists have produced a draft of a four-trait or even an eight-trait rubric in under an hour. Some faculty may need up to ten hours, spread over time.

Why should you spend the time? When Anderson was constructing her scale, she was teaching a twelve-hour load, working for her doctorate at
an institution fifty miles from her home, and was a single mother with two teenagers and a baby. She did it because she wanted to:

- Make grading more consistent and fair.
- Save time in the grading process. Once she was very clear about what she was looking for and had a scoring rubric, she could move quickly through students’ work.
- Diagnose her students’ strengths and weaknesses very specifically in order to teach more effectively.
- Track changes in her students’ performance over several semesters so she could see how changes in her teaching affected student performance (see Anderson and Walvoord, 1990).

Here are some reasons that other faculty in our experience have found it worthwhile to do rubrics:

- To help teaching assistants (TAs) grade papers consistently.
- To reach agreement with colleagues on criteria for common exams, multiple sections, or sequenced courses.
- To introduce greater distinctions into one’s grading. For example, a psychologist had written a set of loosely stated criteria, but she found herself giving A’s to papers she felt did not deserve an A. Somehow she had not captured in her loose list the full range of criteria she wanted to use to make distinctions. A full rubric, with its greater specificity, helped her tease out for herself what those criteria were and then to distinguish the truly excellent papers from the others.
- As data for departmental and general education assessment (see Chapters Eleven and Twelve).

**Using Rubrics for Grading**

Rubric scoring and grading are not the same thing, though grades can be derived from rubrics. Anderson had several reasons for using rubrics, not all of them linked to grading. For example, she could track whether her students improved from one semester to the next on a trait such as “experimental design.” A paper might get a low grade but score high on that trait. In this case, the teacher’s attention would be focused on the design score, not the grade.

Nonetheless, rubrics can lead directly to a grade if the teacher wishes. There are several ways to do this. First, constructing a rubric may act as a clarifying exercise to inform grading in a general way, making the teacher clearer about criteria. In this case, the teacher can continue using a grade description or a checklist and let her own private construction of the rubric inform those two types of documents.
A second strategy is to share the rubric with students but not translate it into a grading instrument. Students use the rubric to guide their work, and the teacher uses the rubric to guide her grading in a general way, but the actual grading is done holistically, without assigning specific points to items in the rubric. The rubric thus works for both students and teacher as a guide, not as a calculator.

A third approach is to translate the rubric scores into a grade. For example, Anderson’s rubric awards a separate score on each of the ten traits for a scientific report. To reach a single grade for the entire student’s work, the traits have to be weighted. The maximum number of points students can gain on each item will be different, depending on the importance the teacher attaches to the item. The total points for all the items combined can then be translated directly into a grade. For example, in grading students’ scientific reports, Anderson used the five “sciencing” sections of the rubric as worth a potential maximum of 20 points each. She multiplied the rubric score by 4, so a rubric score of 5 yielded 20 points; a rubric score of 1 yielded 4 points. She counted the writing skills sections of the rubric as worth a potential of 10 points (she multiplied the actual rubric score by 2). Now she had a total number. She took two-thirds of the total as the project grade.

Some faculty construct a regular rubric, but instead of a scale ranging from 5 to 1 on every item, the rubric assigns points. For example, in rating a speech, the trait of organization might have a scale of 20 points down to 1; eye contact with audience might have a scale of 5 down to 1. After scoring each item, the teacher counts the total number of points. Other faculty designate a percentage for each rubric item. That is, statement of the thesis counts 10 percent of the final grade, quality of evidence 20 percent, and so on.

Frequent Questions in Developing Rubrics

In working with faculty as they construct rubrics, we have found some common questions.

At What Level of Generality Should I State the Traits?

Everyone we have known who composed a rubric has struggled with this issue. Let’s take Anderson’s biology scale as an example (Exhibit 4.6). Her “materials and methods” trait incorporates other traits: how replicable and quantifiable the data are, whether information is properly sequenced, and so on. Some of those might become separate traits, but there is no absolute rule about this; you can state traits at whatever level meets your needs. As your scale develops, you may not only disaggregate traits but also combine them.
Typically traits continue to be revised over time. For example, when Anderson composed her scale, she chose “nonexperimental information” and “control of variables” as separate traits. Subsequent experience showed her that as her students became more proficient at controlling variables (the more inclusive skill), their ability to recognize nonexperimental information also improved. She might, if it were worthwhile to take the time, now fold the two together.

**Which Forms of Language Should I Use?**

We examined a type of rubric that describes the highest level of performance and then simply lists, but does not describe, the levels of performance (Exhibit 4.5). In a full rubric, however, the traits are stated as nouns or noun phrases. This keeps the naming of the trait separate from descriptions of the levels of performance. We think this rule should be broken only if you have thought carefully about it and made a conscious decision to change it for good reasons. Here are traits taken from two faculty members’ drafts of rubrics. They are not stated as nouns:

- Trait from an English course: “Make inferences that are supported by specific information presented in the paper.”
- Trait from a statistics course: “Obtain an appropriate random sample of members of the population.”

If you use verb phrases such as these instead of noun phrases, the traits in essence become a command or a description of actions or of levels of performance. The character of the scale changes, and confusion may result. We suggest changing the traits just mentioned to nouns:

- **English revision**: “Inferences”
- **Statistics revision**: “Random sample [or sampling]”

In a full rubric, the traits are nouns, but the scales within each trait are descriptions of student work. Scoring is done by matching student work to the description within various levels. For example, in a draft scale for a computer spreadsheet project under the trait “spreadsheet accuracy,” the teacher took over some of the language from the assignment sheet: “Spreadsheet must provide . . .” It is better for the teacher to leave out “must” because she is describing rather than giving orders. A revision would read, “[A top-level] spreadsheet provides . . .”

Be careful about ascribing things you cannot know. A draft of a scale for essays in French reads, “Student has made an attempt to follow correct French verb forms and syntax.” It would be better to just describe the verbs
and syntax that will appear in the paper because no professor can know whether the student made an attempt.

How Many Levels Should the Scale Contain?
You may have any number of traits, such as “title” or “materials and methods.” Under each trait, the number of levels depends on your purpose. You could have a two-level scale, which is a pass-fail scale. We have illustrated five levels, but three or four or six are certainly possible. How many distinctions do you need to make? How many can you make, given your insight and the language you can find to describe what you want? Teachers often begin with two or three distinctions and then gradually find ways to distinguish additional levels. For rubrics to be used by multiple scorers, as in departmental or general education assessment, one school of thought holds that four levels are better than five because with five levels, a large percentage of scorers may settle on the middle score of 3.

What Are the Relations Among Levels in a Scale?
The most common strategy for describing performance in a rubric is to begin with the optimum performance as the top value and describe the lower levels of performance as less than, or lacking, the qualities of the top performance. Anderson’s title rubric was modeled in that way. Another is to fully describe the baseline performance and then identify value-added features for the scale. For example, the lowest level contains valuable information about the topic. The next level does that and also contains a thesis or main idea. The third level does those things and also uses the information in a logical way to support the thesis. And so on.

A second approach is for the levels to represent different qualities. Here is a draft from sociologist Grace Auyang of Raymond Walters College of the University of Cincinnati for an assignment where students must respond to sociological articles:

Trait: Approach to the Problem Described in the Article

5 Student synthesizes the problem.
4 Student analyzes the problem.
3 Student explains the problem.
2 Student describes the problem.
1 Student merely identifies the problem or does not address the problem.

Auyang provides further descriptions of what she means by the terms identify, synthesize, and so on, but our point here is to show in skeletal form that the scale is not quantitative; each level represents a different skill.
How Should I Handle Concepts such as “Adequate” and “Appropriate”?  
You can use such terms as long as they are specific enough for your needs. If you want to explain your criteria to students, ask yourself (and them) whether they know what “appropriate” means or whether you need to be specific. If you are preparing the rubric scale for use by an outsider, ask whether that person is sufficiently familiar with your discipline to know what you mean.

Let us look at the use of appropriate, correct, and similar terms in an example. Here is the draft of Bill Marsh’s rubric scale for a statistics assignment at Raymond Walters College of the University of Cincinnati. In this study, students construct a null hypothesis, select a random sample, collect data, and draw conclusions. Here is a portion of his rubric (his full rubric is example 9 in Appendix A). Note his use of words such as correct and appropriate.

Trait: Methodology

5 Correct statement of problem with accompanying null and alternative hypothesis. Well-defined population with appropriate random sample. Data collection is free of bias or contamination.

In this discipline, statistics, the words correct, appropriate, and so on refer to commonly accepted procedures on which professionals generally agree. It would take pages of a statistics textbook to explain what an “appropriate” random sample would be for each of the projects students might choose. Marsh will spend the semester teaching his students the meaning of correct and appropriate in these settings. Whether and to what extent the teacher would enlarge the rubric to explain these terms would depend on the audience and purpose.

A different slant on words such as correct and appropriate is presented by a historian, Jim Cebula of Raymond Walters College, who in this rubric draft is trying to describe what he calls the elegance of argument for history essays:

Trait: Elegance of Argument

5 Original and clearly stated thesis; persuasive, well-organized, imaginative use of source material
4 Clearly stated thesis, good use of sources, well organized
3 Facts straight with a reasonable explanation of the subject under consideration
2 Poorly stated thesis, inadequate survey of available sources, poor organization
1 No awareness of argument or complexity
This rubric contains words of quality such as good, well, and poor. Depending on audience and purpose, the faculty member might want to define these. There are also words that describe characteristics, such as original, persuasive, and imaginative. Again, depending on audience and purpose, the teacher may want to try to pin down at least some of the characteristics by which she or he decides that a student's work is original or imaginative.

When teachers are trying to construct rubrics, words such as those we have examined here crop up regularly. Decisions about such terms must be based on use. Is this teacher simply trying to articulate her criteria for herself and her students? If so, whatever level of specificity is needed for her own and her students' understanding will be enough. If these rubrics are being used for departmental or institutional assessment, some terms may need further definition, depending on how much agreement there is among the external readers about what correct in statistics or original in history means.

When dealing with such issues, do not forget the power of examples. Here is part of a primary trait scale draft by Ruth Benander of Raymond Walters College for a first-year composition course. Her assignment calls for students to read several essays on a single topic and then write an essay that synthesizes the readings. Here is her scale for the trait “Synthesis of Ideas”:

**Trait: Synthesis of Ideas**

4 Presents a perspective that synthesizes the main ideas of several readings. This perspective creates an informative way to view the several main ideas of the readings in a way that gives more meaning to the readings as a whole rather than if the main ideas were presented individually.

Example: “Urban and rural violence may differ in frequency but not in the intensity with which they affect the lives of the people involved.”

3 Presents a perspective that synthesizes the main ideas of several readings. This perspective may be very general.

Example: Violence is everywhere and affects us all.

At least two different readings are presented, though they may not be clearly related to each other under the umbrella of the synthesizing idea. The two different readings chosen may demonstrate similar rather than different views of the perspective that synthesizes them.

2 The main idea of one reading is presented as the dominant perspective of the paper.

Example: “The article ‘Gangster Wake-Up Call’ deals with gang violence.”

The main points of the reading are used to support the main ideas of the reading. (The paper will look more like a summary of the article than a synthesis of ideas.)

No alternative views are presented.

1 There is no clear main idea to the paper.

A reading may be discussed, but the main idea is not related to any other ideas.
Such examples are one way to clarify broad or vague concepts within a rubric. Another way is to focus on describing those physical, observable, or even, if possible, quantifiable aspects that led you to decide what is creative or appropriate to the conventions of the discipline. For example, an education department was struggling with how to reach consensus on evaluating students' senior portfolios. The portfolios contained three kinds of material: lesson plans, observations of individual students or of classrooms, and essays in which the seniors considered theoretical or practical issues in teaching and learning.

One element that all of the professors valued as a trait for the portfolios is what they called reflective practice. It means that the student teacher habitually reflects on and theorizes about his or her teaching experience rather than merely following unexamined practices or recording events.

To try to describe reflective practice more concretely, they took some portfolios they all agreed were outstanding in reflective practice and some they agreed were poor in that quality. Then they said, “There must have been some physical, observable characteristics in these written artifacts that led us to conclude that this portfolio shows a reflective practitioner and this one does not. What were they?” Carefully examining the students’ texts, the faculty saw that the students who scored high on reflective practice shared these characteristics:

- They routinely stated why they themselves, or students with whom they were working, exhibited certain behaviors.
- They explained reasons and outcomes in terms of theoretical statements or hypotheses about teaching, usually expressed in the present tense.
- They routinely recorded questions and dilemmas.
- They recorded and described patterns in their own or in others’ behavior.
- They reported having done reflective thinking. (“All that week, I thought about why Randy’s behavior had surprised me so much.”)
- They exhibited a musing, questioning tone.
- They noted missing information needed to draw conclusions from data or experience.
- They hypothesized as to causes, connections, or outcomes.

Detailed descriptions of the physical characteristics of the student’s work can help you be specific about qualities that are indicators for abstract terms such as reflective practice.
What Kinds of Student Performances Can Be Scored by Rubrics?

Almost any type of student performance—oral, clinical, artistic—involving higher-order thinking, creativity, or integration of skills can be examined effectively with a rubric. For example, Judith Bloomer of the Department of Occupational Therapy and Evelyn Lutz of the Department of Nursing at Xavier College in Cincinnati assessed these traits in evaluating their students in Work-Related Interactions with Others: Collaboration, Participation, Attitude, Independence, Communication, and Responsiveness. (See the full scale in example 2 in Appendix A.)

Team performances can also be examined effectively with rubrics. Lawrence Fredendall, at Clemson University, developed a twenty-four-trait rubric to help members of local businesses evaluate teams of student interns. The rubric includes criteria for group behaviors such as team punctuality, courtesy, appearance, communications, and enthusiasm. (See example 7 in Appendix A.)

What If the Scale Leads Me to a Score That Does Not Feel Right?

Suppose that you find yourself giving high scores to student work you find competent but somehow lacking in the originality, creativity, or risk taking you want for A work. Often the problem is that you have not included in the scale all the traits you are using. Look again at the student work that makes you uneasy. Compare it with samples that score above and below it. Ask yourself these questions: What is missing in my scale? What is most important? Try to capture that in the scale. Use the best language you can. It is fine to say, as one faculty member did, “This paper shows a curious mind at work.” If you don’t want to give an A to any work that does not have originality, then weight your scale accordingly.

How About Involving Students in Constructing Standards and Criteria?

We know faculty who like to involve their students in establishing criteria for student work. For example, a physicist whose students give oral reports on a physics project asks the students to brainstorm a list of the qualities that make an effective oral presentation from their point of view. Students typically offer items such as “speaker talks clearly and can be heard by everyone,” “has good eye contact,” “uses charts and graphs that are clear and correctly labeled,” “explains the experiment in a well-organized way,” and so on.
The instructor can then incorporate these when building the rubric scale. For example, some traits suggested by these sample student comments might include "delivery," "charts and graphs," and "organization." In composing the scale under "delivery," the students' comments suggest that a high score might include speaking clearly, being heard by everyone, and maintaining consistent eye contact with all areas of the room.

**Can Rubrics Be Used with TAs as Graders?**

Yes, very effectively. Gisela Escoe, Philip Way, and their graduate assistant, Jack Julian, teaching economics at the University of Cincinnati, used rubrics to help TA graders in large introductory economics classes. They held an introductory session in which the TAs went over the rubrics and scored sample papers. During the semester, the TAs graded all the papers. They were encouraged to bring to the faculty any papers about which they are unsure.

When TAs are new or inexperienced or when TA populations turn over frequently, a common practice is for a single faculty member or small number of key personnel to construct the rubric. Faculty then hold a TA training session on using rubrics. However, in a team-taught course, multi-section courses, or a situation where TAs participate as colleagues in establishing criteria and standards, the goal is to establish a collaborative way of arriving at rubrics. In order to maximize participation while not wasting time, we suggest following a procedure.

In step 1, your group establishes course goals, key assignments, and, the criteria and standards by which the most important assignments and exams will be evaluated and by which students will receive ongoing feedback on their progress.

Working from the basic outlines established in those discussions, the group next meets to establish the rubric traits and scale to be used for each major assignment and test.

If you have sample papers, you might begin by talking about which grade you would give them and why. This discussion will reveal traits. For example, suppose someone says, "I'd give this paper a B because it has a clear problem statement and it analyzes the problem well, but the solutions do not take into account all of the important factors." At least three traits are embedded in that statement: (1) the problem statement, (2) analysis of the problem, and (3) solution. Someone should be listing these traits on a board or newsprint pad. Do not try to assign weights to the traits at this point (though you might want to jot down information that emerges about the relative importance of traits for later use). If the group disagrees, try to pinpoint the traits that are at issue and to reach consensus.
You will have traits stated at various levels of generality. For example, in this situation, the "solution" might be a trait embodying several considerations including the factors, but "factors considered in the solution" might also be a separate trait with its own scale. At this stage, jot down people's insights about which traits are overlapping or subordinate, but do not try to make decisions about aggregating or separating traits in the large group because it will take too long and lead to confusion.

As the discussion proceeds, it may emerge that the assignment or even the entire course plan needs to be changed or explained more clearly. If you have time, let that happen. One of the most useful outcomes of discussions about standards and criteria is the opportunity to rethink course goals, course structure, and the nature of tests and assignments.

In step 3, the group should decide how many levels the rubric needs at the start. We suggest three to six levels.

Now, in step 4, we suggest that one or two people be tasked with drafting a rubric using this brainstormed list. When they prepare and distribute the draft, the entire group offers suggestions for changes in it, which the writers then incorporate. Repeat this cycle until everyone is comfortable with the scale (or until everyone is too tired to object). If you cannot reach consensus on all the traits, consider the possibility that common traits on which people do agree may be used by everyone and give individuals the freedom to construct their own scales around other traits.

As soon as you have sample student papers, step 5 is to let the group score papers using the scale. Have individuals score papers independently at first, meet to compare scores, discuss discrepancies, and change the scales as needed.

In the final, and ongoing, step, you will hold periodic discussions which sample papers are scored, discrepancies discussed, and the scale and assignments changed as needed.

This procedure assumes everyone's participation. If there are too many people for this approach, a committee may do this work. A useful description of how composition faculty from six different institutions together constructed a rubric for student work can be found in Pagano and others (2008).

**Can Rubrics Be Used for Portfolios?**

Yes. We define a portfolio as a collection of work by the same student completed over time. Your department or assessment committee can follow the same basic process we have just outlined. The reason for using a portfolio rather than a single piece of work from each student is to evaluate attributes that can be assessed only by multiple works—for example, the consistency
of a particular student performance in multiple situations, a broad range of student skills, the flexibility with which the student applies principles to varying situations, or the growth of the student’s skills or knowledge over time. (Chapter Eleven provides a discussion of using portfolios as part of departmental or general education assessment.)

**Can Rubrics Be Used for Multiple-Choice and Short-Answer Tests?**

Yes. Rubrics are particularly useful for these tests because they force the teacher to state explicitly what skills and knowledge she expects and to provide students with direct feedback about how well they are achieving course goals. A process that has been called test blueprinting shows how to connect course goals, test questions, and rubrics (Middle States Commission on Higher Education, 2007; Suskie, 2009). For example, Table 2.1 shows how a teacher might generate test questions linked to rubric traits such as “reading graphic information” or “making connections between graphic information and scientific issues.” Table 4.1 shows how you might go further, filling out the full rubric, including descriptions of the levels of performance, and then matching students’ test answers to levels of the rubric. Table 4.1 contains one trait; the full rubric is in example 14 in Appendix A.

When you draw test items from a test bank, perhaps provided by the textbook publisher, you may need to consider not only what type of reasoning the test item calls for, but whether it asks merely for recall of textbook material or requires students to reason on their own. For example, Patricia Schlecht, of the Nursing Department at Raymond Walters College of the University of Cincinnati, categorized the multiple-choice questions

---

**TABLE 4.1**

Students’ Test Answers Mapped to Rubric Levels

<table>
<thead>
<tr>
<th>Trait: Reading Graphic Information</th>
<th>Percentage of Students Answering at This Level on Relevant Sets of Multiple-Choice, True-False, and Short-Answer Questions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Level 4. Accurately describes the function of the graph; identifies points and ranges</td>
<td>33 percent</td>
</tr>
<tr>
<td>Level 3. Partially identifies the functions of the graphic; identifies points and ranges</td>
<td>55</td>
</tr>
<tr>
<td>Level 2. Identifies points OR ranges only</td>
<td>12</td>
</tr>
<tr>
<td>Level 1. Lacks competency</td>
<td>0</td>
</tr>
</tbody>
</table>
that she gave students on tests by deciding whether the question required a high level of thinking, using Bloom’s *Taxonomy of Educational Objectives* (1956), and whether the textbook or her lectures had given students the answer directly. She pointed out that although a multiple-choice question may appear, by Bloom’s taxonomy, to require critical thinking, it may in fact only ask the student to repeat material given in the textbook or in lecture. She developed these categories:

- **Higher critical thinking.** Questions would fall in the analysis, synthesis, or evaluation levels of Bloom’s taxonomy. Course materials give needed background to answer the questions. There is no directly visible connection between the course material and the test question.
- **Lower critical thinking.** Questions would fall in the application level of Bloom’s taxonomy. Course materials give needed background to answer the questions. There is a directly visible connection between course material and the test questions.
- **Knowledge and comprehension levels of Bloom’s taxonomy.** Material is directly from the course presentation, with some changes in wording and phraseology.

Using such well-defined levels to categorize critical thinking, faculty can group multiple-choice items into critical thinking sets. Exhibit 4.9 is a hypothetical rubric.

More important, this classroom assessment strategy has the potential to collect aggregated student data that can inform and improve the whole nursing program (see Chapters Eleven and Twelve for more examples of how to use rubrics and test blueprinting to collect valuable data from multiple-choice and short-answer questions).

**EXHIBIT 4.9**

**Rubric for Critical Thinking Performance on a Twenty-Five-Item Multiple-Choice Test**

<table>
<thead>
<tr>
<th>Score</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>Answers correctly all or all but one of questions 4, 7, 12, 13, 20, 21, and 24 (designated higher critical thinking questions on a given test)</td>
</tr>
<tr>
<td>3</td>
<td>Answers correctly 4 or 5 of the higher critical thinking questions. Total test score is 70 percent or above.</td>
</tr>
<tr>
<td>2</td>
<td>Limited evidence of critical thinking (only 1 to 3 questions correct within the critical thinking set); however, score for total test score is 70 percent or above.</td>
</tr>
<tr>
<td>1</td>
<td>Lacks evidence of critical thinking and/or score for total test is 68 percent or below.</td>
</tr>
</tbody>
</table>
Selecting a format—a grade description, a checklist, or a rubric—to express your standards and criteria is an important cornerstone for your teaching. When well constructed, such instruments show you what to teach and help students learn the standards for good work in their field.

**Criteria Outside the Rubric**

Rubric scores can be used to establish only a portion of the grade. Anderson reserved a small portion of the grade to evaluate whether the student had followed the word limit, met deadlines for rough draft submission, and conducted the required peer editing conference. Dorothy Solé’s scale (Exhibit 4.1) determines only a portion of the grade given to student journals in a beginning Spanish class. The other portion is based on length and number of entries.

**Gateway Criteria**

The instructor can establish what we call gateway criteria in which students are asked to comply with certain requirements before the paper is even subjected to rubric scoring. In other words, students must meet certain requirements even to get into the ballpark where the rubric score will be used. Work that does not meet the gateway criteria is returned to the student with a failing grade or instructions to revise and resubmit. So if a student turns in a final scientific report that does not meet the teacher’s announced standard for word processing, labeling of graphs, grammar, punctuation, or other aspects, the teacher responds by handing it back with either an F grade or instructions to revise and resubmit for grading. Such a report does not make it into the ballpark, so it does not even get to the point of applying a rubric.

**Criteria for Grammar and Punctuation**

Issues of grammar and punctuation can be integrated into the grading scheme in several ways.

**Grammar and Punctuation as Gateway Criteria**

One strategy is to establish a certain level of grammar and punctuation as a gateway requirement. The philosophy behind this approach is that if a student later submits a proposal to her boss or a letter to a client that does not adhere to standard English grammar and punctuation, the boss or client may simply dismiss the writing. The truth is that societies do use language as a sorting device, making judgments about writers based on their command of the standard language.
We suggest that you not only explain to students your policy about use of standard English, but that you also establish a research-based understanding of language in your classroom. Exhibit 4.10 is a sheet that

**EXHIBIT 4.10**

**Handout Used to Explain Walvoord’s Gateway Criteria**

**Policy for Use of Edited Standard Written English**

Suppose a group of people were living on an island, all using the same language, until one day the island broke in two, separated by impassable water. In one hundred years, with no contact, would the people on both halves still use the same language forms? No. Human language is always changing. Language on each half of the island would evolve with different forms and rules. Neither would be better in any absolute sense—just different. Similarly, in the United States, language variations have developed among people separated by culture, socioeconomic status, ethnic background, or geography.

Nevertheless, the language of the ruling class commonly comes to be regarded as standard. In the United States, the “standard” is the language of the white middle and upper classes. Forms of English developed by people of color and by people who have been poor or geographically isolated (as in Appalachia) are sometimes said to be “bad” or “incorrect” English, but such forms are only different, not bad. Each form of English has its own rules. People who say “she working” are not speaking “bad” English; they are using a different set of rules for forming the present tense.

One of the tasks of a good education is to make students aware of these facts about language. Another task of education, however, is to prepare students to function effectively in the world where readers generally expect writers to use edited standard written English (ESWE). Thus, in this class, you must use ESWE. Here is the standard I will apply.

On finished, final, formal papers (not on drafts, in-class writings, or writing that I specifically label as informal), you must have no more than an average of two departures from ESWE per page, in any combination of the following areas:

- End-of-sentence punctuation (avoid run-on sentences, comma splices, fragments, or misuse of semicolon). Occasionally you may use a fragment or comma splice for a special effect. Label it in the margin.
- Verb forms (use ESWE rules for adding -ed and -s, for using helping verbs, and so on).
- Verb tense (avoid confusing shifts in verb tenses).
- Agreement of subject and verb.
- Pronoun form (use ESWE rules to choose between I and me, she and her, who and whom, and so on).
- Agreement of pronoun with antecedent (the antecedent is the word the pronoun refers to).
- Use of apostrophe s and the -es.
- Use of quotation marks for all quoted words.
- Spelling (a typo counts as a misspelling).
- Proper sentence sense (no words omitted, scrambled, or incomprehensible).
Walvoord used to explain to her students a gateway policy of grammar, punctuation, and spelling and a research-based concept of how standard and nonstandard languages function in a society. Note that Walvoord did not apply this policy to drafts, in-class writing, or informal writing, only to finished, formal work.

Walvoord reminded her students that the writing center would help them with these criteria, and she gave students a list of the writing center’s Web page, hours, and location. She reminded her students about her gateway policy frequently, both orally and in writing, and let her department head know that she was implementing this policy. She offered students an opportunity to bring her, or send electronically, a typed draft of the paper at least twenty-four hours before the paper was due. She did not edit these papers by finding all the departures from edited standard written English for the students (finding their departures from standard English was their job), but she made a judgment about whether the paper met the gateway criteria. If it did not, she offered suggestions about writing center help and about the most common problems she saw, so students could look up those items in a composition handbook or online. In other words, she tried to integrate the real-world criteria that students must meet with the support and help appropriate to a learning environment.

The result of this gateway policy is that virtually all the final papers Walvoord receives make it through the gate. Walvoord has used the policy with first-year students in a selective private college and at a state research university.

You may want to set the criteria at different levels for different groups of students. You will also want to consider the effect of this policy on students who speak English as a nonnative language. You can hold nonnative speakers to the same standard, asking them to get help at the writing center or the English language center, or you can establish separate levels for them that stretch them beyond where they are but do not present them with an impossible task. The idea is not to hand out a lot of F grades but to teach students that to function in the outside world, they will have to master ESWE or their work will be dismissed before the reader has even dealt with the writer’s ideas.

**Integrating Grammar and Punctuation into the Grading Scale**

Instead of a gateway policy, some faculty integrate grammar and punctuation into the grading scheme. For example, Breihan, the historian whose grade-based sheet appears in Exhibit 4.2, integrates grammar and
punctuation into the sheet. One factor in the description of a failing essay is to be “incomprehensible due to errors in language or usage.” One factor in a C paper is that meaning is “all or partly obscured by errors in language or usage.” The advantage of this mode is that, as in real life, the management of grammar and punctuation is closely intertwined with other aspects of the student’s writing, and it combines with those aspects to make an impression on its reader.

Other faculty members integrate grammar and punctuation into their check sheets. For example, a check sheet may ask students to check off that, before submitting the paper, they have used spellcheck, have reread the paper at least twice for grammar and punctuation, and have asked one other person to read the paper to check for grammar and punctuation.

Many faculty construct a separate item in the rubric for grammar and punctuation—for example:

5 No departures from ESWE.
4 Occasional departures from ESWE, but not in the critical areas listed in Exhibit 4.10.
3 No more than two departures from ESWE in the critical areas.
2 More than two departures in critical areas, but writing is still readable.
1 Departures from ESWE are very frequent, interfering with smooth reading.

A developmental approach is to have each student keep a written record of his or her most common problems with grammar and punctuation. For example, Amy, based on her earlier papers or her own self-knowledge, might list use of the apostrophe as one of her common problems. The check sheet then asks students to assert that they have checked the paper especially carefully for their own issues. You can ask each student to establish a “never again” list. With each paper, the student adds to the list one area of ESWE that she vows “never again.” This sheet is handed in with the final paper, and the writer is held responsible for the “never again” list, but not for other aspects of ESWE. This approach can be very helpful for nonnative speakers of English. An instructor can work with a writing center or English language instructional program to define the course of a student’s progress and the ESWE issues for which that student will be held responsible.
<table>
<thead>
<tr>
<th><strong>Activity</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>If you have been following our course planning procedures, you now have a course skeleton, a draft of at least one assignment, and a big-picture view of linking criteria and grading. Now we invite you to do the following:</td>
</tr>
<tr>
<td>- Construct a list of traits for one of your assignments or tests. You may simply choose two or three traits that seem most important, or you may try to construct a list of all the traits you want to use for grading. Remember that some aspects, such as editing for edited standard written English or avoiding numerical errors, may be handled as gateway characteristics separate from the rubric scale.</td>
</tr>
<tr>
<td>- Construct a grade-based guide, a check sheet, or a rubric for your traits. For the rubric, you may use a two-, three-, four-, or five-level scale, depending on how many levels you can or want to construct.</td>
</tr>
<tr>
<td>- Score or grade one or several assignments or tests with your instrument, and revise it as needed.</td>
</tr>
<tr>
<td>- Revise your draft of assignment instructions to students to reflect clearly your criteria and standards. Remember that you may include a grading description, a check sheet, a rubric, or other language that makes criteria and standards clear.</td>
</tr>
</tbody>
</table>