



Guides to Scoring Student Work: Checklists and Rubrics is an excerpt from *Put to the Test: Tools & Techniques for Classroom Assessment* by Therese M. Kuhs, Robert L. Johnson, Susan A. Agruso, and Diane M. Monrad.

© 2001 by Therese M. Kuhs, Robert L. Johnson, Susan A. Agruso, and Diane M. Monrad.

All rights reserved. No part of this material from *Put to the Test* may be reproduced in any form or by electronic or mechanical means, including information storage and retrieval systems, without permission in writing from the publisher, except by a reviewer, who may quote brief passages in a review.

4 | GUIDES TO SCORING STUDENT WORK

Checklists and Rubrics

Anytime teachers use open-ended assessments that require more than a simple right or wrong answer, students might question the subjectivity of grading. Teachers can preempt such concerns by developing written statements that will guide the evaluation of student work. Such written descriptions, called scoring guides, are useful tools because they outline what teachers expect to see in student responses. These expectations list important knowledge and skills to be used as evaluative criteria in grading students' work.

A checklist is the simplest form of scoring guide. In Chapter 2, we focused on the use of checklists for recording information from an observation. In this chapter, our discussion will address checklists designed to evaluate student products. In both cases the checklist outlines evaluative criteria for assessing student learning. Beside each criterion is a space to indicate whether that criterion has been met or not. Figure 4–1 is a checklist for evaluating limericks written by third-grade students. It contains the important elements that students studied as characteristic of a well-written limerick. Notice the place in front of each criterion that can be checked if the student's limerick shows evidence of proficiency.

Clearly there is little question about the first criterion: A student's limerick either has five lines or it doesn't. Some of the other criteria, however, might be met to different degrees. For example, what if a student had one

Student Name:	Date:
<input type="checkbox"/> Has five lines.	<input type="checkbox"/> Has correct capitalization.
<input type="checkbox"/> Tells a funny story.	<input type="checkbox"/> Has correct punctuation.
<input type="checkbox"/> Has correct rhyming pattern.	<input type="checkbox"/> Contains descriptive words.

capitalization error, had forgotten one comma, or had used only one descriptive word? Should she get credit or not?

Perhaps the teacher would want to award a different number of points to reflect the different levels of proficiency students might demonstrate. To do this, the checklist could be revised to be a different kind of scoring tool, a point-allocated checklist, as shown in Figure 4–2. A point-allocated checklist indicates the number of points for each criterion. In this example, one point would be awarded for the first criterion, but up to three points could be earned for each of the other criteria.

Using this point-allocated checklist allows the teacher to award partial credit depending on the student’s level of proficiency. The levels of proficiency are implicit, rather than explicit because the list does not provide descriptions for how one, two, or three points are to be awarded. The teacher must decide, for example, how many and what kind of descriptive words merit a score of 3 rather than 2. Yet the student is unaware of the teacher’s implicit rule. Thus, when using this type of point-allocated checklist, teachers frequently find themselves justifying their decisions to students about how many points they awarded.

The remedy to this situation is a scoring tool that combines evaluative criteria with scales that explicitly define standards of performance. This type of tool is called a rubric. Figure 4–3 illustrates the detail needed to turn the point-allocated checklist in Figure 4–2 into a rubric. In this case the scoring guide is called an analytic rubric because there are separate scales for each evaluative criterion.

This example illustrates many of the things teachers must consider in developing a rubric. Not only must the teacher select the evaluative criteria,

Student Name:	Date:
1 point	Score
Has five lines.	_____ point
3 points each	
Tells a funny story.	_____ point(s)
Has correct rhyming pattern.	_____ point(s)
Has correct capitalization.	_____ point(s)
Has correct punctuation.	_____ point(s)
Contains descriptive words.	_____ point(s)
	_____ Total

FIGURE 4–2 Point-allocated checklist for limericks

Student Name:	Date:		
Criteria	Scales		
Has five lines.	0 Limerick does not have 5 lines.	1 Limerick has 5 lines.	
Tells a funny story.	1 The story that the limerick tells is incomplete <i>and</i> is not funny.	2 The story that the limerick tells is incomplete <i>or</i> is not funny.	3 The story that the limerick tells is complete <i>and</i> funny.
Has correct rhyming pattern.	1 Limerick does not rhyme.	2 Limerick has some rhymes, but the rhyming pattern is not correct.	3 Limerick has correct rhyming pattern (AABBA).
Has correct capitalization.	1 Capitals are not used or are used incorrectly.	2 Capitals are sometimes correctly used.	3 Capitals are correctly used in the poem and title.
Has correct punctuation.	1 Limerick is not punctuated or is punctuated incorrectly.	2 Limerick has some correct punctuation.	3 Limerick is correctly punctuated.
Contains descriptive words.	1 Specific nouns, adjectives, and adverbs to paint a picture for the reader are not used.	2 The selection of specific nouns, adjectives, and adverbs to paint a picture for the reader is attempted.	3 Specific nouns, adjectives, and adverbs to paint a picture for the reader are effectively selected.

FIGURE 4-3 Analytic rubric for scoring limericks

she must also develop a scale. In developing the scale, the teacher must decide how many levels of performance to define and write descriptions that clearly and appropriately outline expectations for the different levels of performance. By combining numbers and descriptions, the teacher explicitly defines standards of performance that she will use in grading student work.

As was the case in the point-allocated checklist (Figure 4–2), two different scales are used in this analytic rubric. A scale with two levels (0 or 1) is used to indicate whether the student’s work followed the convention of a limerick having five lines. For the other evaluative criteria, three-point scales describe levels of student proficiency. Also note that the descriptions of each criterion address the same elements at each proficiency level to define what would be excellent, satisfactory, or unsatisfactory performance. For example, the description for “Tells a funny story” talks about completeness and humor at each level of performance. We will use the term *parallel structure* to refer to the practice of repeating the elements of performance across all proficiency levels of a criterion.

In an analytic approach, several scores are given for each piece of work. If the scores are to provide teachers, students, or parents with feedback about specific skills that need improvement, then a checklist or analytic rubric is very useful. For example, the scores from the analytic rubric for the limerick could tell one student that she has strength in the use of descriptive words but still needs to work on punctuation while telling another student that he needs to use the correct rhyming pattern but his limerick is punctuated correctly. A teacher can use this detailed information in planning future lessons for the class or particular experiences for individuals.

If the purpose of the evaluation is to provide a description of the student’s performance at the end of a unit of study, a holistic approach to scoring is sometimes useful. Figure 4–4 illustrates a holistic rubric that could also be used to evaluate students’ limericks. All the criteria addressed in the checklist (Figure 4–1) and the analytic rubric (Figure 4–3) are also addressed in this holistic rubric. The difference is that, when using the holistic rubric, the teacher must make a single judgment—that is, award a single score—about student learning while considering all of the different criteria that she is evaluating. For this reason some teachers find the use of holistic scoring more difficult than an analytic approach.

The holistic rubric also illustrates the use of proficiency labels for each level of the scale: Expert, Accomplished, and Novice. Some teachers use only the labels, others use a combination of numbers and labels, and still others use only a number or score for each level. When numbers are used, higher numbers should indicate better levels of performance to facilitate grading and record keeping.

Holistic rubrics are very useful when it is difficult to separate evaluative criteria in scoring a student response. For some tasks it is better to consider criteria in combination because they are dependent on one another. For example, students might be asked to write a fictional story. In grading student work, the teacher would want to consider if students’ writings included the major components of a story: character, setting, conflict, climax, and resolution. The potential evaluative criteria of conflict, climax, and resolution are intertwined. If an analytic rubric were used, the teacher would have difficulty scoring for the presence of a resolution when a student paper doesn’t

Proficiency Levels	Descriptions
Expert	The limerick tells a complete and funny story. Specific nouns, adjectives, and adverbs paint a picture for the reader. The limerick has the correct number of lines and rhyming pattern. Punctuation and capitalization are correct.
Accomplished	The limerick is characterized by <i>most</i> of the following statements. It may tell a story that is either complete or funny. The use of some specific nouns, adjectives, and adverbs create a picture for the reader. The limerick has either the correct number of lines or the correct rhyming pattern. There are few punctuation and capitalization errors.
Novice	The limerick needs to be written in a way that tells a complete and funny story. To paint a picture for the reader, the limerick needs to contain specific nouns, adjectives, and adverbs. The limerick needs to have five lines and a rhyming pattern of AABBA. Punctuation and capitalization need correction.

FIGURE 4-4 *Holistic rubric for scoring limericks*

really present a conflict. A holistic rubric, like the one in Figure 4-5, would simplify the evaluation.

Up to this point the example checklists and rubrics have been rather specific. The scoring tools are useful either to evaluate student ability to write a limerick or to write fiction. Rubrics can be written for a specific task, as in the previous examples, or can be designed to evaluate a broader range of student work. The latter type of rubric is often referred to as generic. For example, in writing, many experts advocate the evaluation of all writing genres using the criteria of style, conventions, and organization. A teacher could develop a generic rubric to evaluate different forms of writing considering these criteria.

In mathematics there are also certain skills evidenced in good work whether the content of the task is related to geometry, number concepts, or some other area of mathematics. The ability to communicate in mathematics is one area where a generic rubric might be used. Figure 4-6 is a generic rubric that a teacher could use to evaluate a student's ability to communicate in mathematics regardless of the specific task being performed. Readers may want to use this rubric to evaluate student work on the geometry task in Chapter 3 (Figure 3-8). This communication rubric is holistic but the ana-

Proficiency Levels	Description
Level 5	The main elements of a fictional story are present. Descriptive language establishes the setting. The main character is well developed and faces a conflict. The story comes to a climax, and the resolution of the conflict flows from the story.
Level 4	The main elements are present; however, either character or setting is not well developed. The main character faces a conflict. The story comes to a climax, and the resolution of the conflict flows from the story.
Level 3	The main elements are present but not well developed. Characters, setting, conflict, climax, and resolution need elaboration.
Level 2	Some of the major elements of the story are missing.
Level 1	Most of the major elements are missing.

FIGURE 4-5 *Holistic rubric for evaluating fiction writing*

<p>Accomplished (3)</p> <ul style="list-style-type: none"> • Consistently uses mathematics terms and symbols with few inaccuracies. • Presents comprehensive and clear explanations using examples or illustrations when appropriate. • Offers discussion in an organized and logical sequence.
<p>Developing (2)</p> <ul style="list-style-type: none"> • Uses math terms and symbols most of the time but some are inaccurate or incorrect. • Offers correct explanations but statements are unclear or incomplete and examples are either missing or are incorrect. • Responses may be somewhat disorganized.
<p>Not Yet (1)</p> <ul style="list-style-type: none"> • Uses terms and symbols incorrectly or relies on everyday words instead of math terminology. • Explanations and examples are incorrect or not relevant. • Order of discussion may be confusing.

FIGURE 4-6 *Generic rubric for communication in mathematics*

lytic approach can also be used for generic rubrics. Notice the parallel structure of the rubric in Figure 4–6. Each element of performance that is to be considered is mentioned in the description of every level of the rubric.

In contrast, a task-specific rubric can be used only to evaluate responses to a single specific task or project that students are doing. Such rubrics are often useful to evaluate major projects that are done over a period of time or tasks that assess many dimensions of student learning. Figure 4–7 offers a mathematics task with a task-specific rubric that would be used to evaluate student responses to the task. This is an analytic rubric, but the holistic ap-

<p>Activity: We have a record of daily rainfall throughout the year. Use this information to calculate the average rainfall for each month. Create a bar graph to show the average amount of rainfall for May, June, July, August, and September. Begin the graph with May and end with September. Answer the following questions based on your graph:</p> <p>Which month had the most rain? Which month had the least rain?</p> <p>Our city's water department wants to predict when people will be watering their lawns and gardens. Using the information from your graph, tell which month you think people will probably need to water their lawns and explain your prediction.</p>		
Organization of Thoughts	Accuracy of Calculations	Presentation of Bar Graph
3: Answers the questions appropriately and supports predictions based on the graph.	3: Calculations are based on the daily information for each month and are accurate.	3: Bar graph is organized with a title and legends, the data are sequenced by months, and the length of bars is correct.
2: The answers and predictions are consistent with the information in the bar graph, but predictions are not clear or correct.	2: Calculations are based on the daily information for each month and most calculations are accurate.	2: Bar graph has a few mistakes in title, legends, sequence of months, or length of bars.
1: Some answers and predictions are incorrect or are not related to the information in the bar graph.	1: Calculations are based on the daily information for each month. Need to continue to work on accuracy.	1: Bar graph has several mistakes in title, legend, sequence of months, or length of bars.
0: Does not write any statements based on the bar graph.	0: No attempt made.	0: No attempt made.

FIGURE 4–7 Performance task and task-specific rubric

proach could also be used. Although the evaluative criteria—organization of thoughts, accuracy of calculations, presentation of a bar graph—could apply to many tasks done in a mathematics classroom, the elements used in the descriptors refer specifically to requirements of the rainfall task, for example, checking to see if calculations are accurate for each month.

PURPOSE OF RUBRICS AND CHECKLISTS

Rubrics and checklists are powerful tools that help teachers improve assessment of learning. One way that these scoring guides improve assessment is by ensuring that each student's paper will be judged using the same criteria. The explicit statement of expectations for student learning provided in the checklists and rubrics dealing with limericks reminds the teacher to look at capitalization, punctuation, and the rhyming pattern on each student's limerick.

A rubric that explicitly outlines the content that teachers are to focus on supports a teacher's consistency in scoring. Teachers have all experienced the frustration of wondering whether they graded the twenty-fifth student paper in the same way they graded the first paper. The scale in a rubric guides the teacher to award the same number of points to papers that have the same strength or flaw, whether it's the first or the last paper graded.

A rubric also helps a teacher sustain consistent attention to the evaluative criteria. Without a rubric, a teacher may be tempted to give a very low grade to a piece of writing that is riddled with spelling or mechanics problems even though the content and organization of the paper surpasses that of other essays being graded. Use of a rubric reminds the teacher of the relative importance of each criterion and limits the number of points that a student may lose for any one type of error.

Rubrics can also have instructional implications. Ideally, the teacher should develop a rubric before instruction begins. Imagine a teacher sitting down to plan a week's lessons for language arts, realizing that the work students will submit on Friday will be evaluated using the analytic rubric for limericks (Figure 4–3). In preparation for that week's classes, the teacher would plan activities to emphasize ideas about a limerick's length, story line, rhyming pattern, capitalization, punctuation, and descriptive words. In this situation the rubric can function as a quick summary of important knowledge and skills to be taught during the week and can help the teacher plan lessons that target important content.

The reality of most classrooms, however, is that teachers usually develop the assessments when the unit is nearly finished. When teaching the same content in subsequent years, the teacher will have last year's rubric to support planning. By reviewing or designing the rubric prior to giving students an assessment task, the teacher can determine if she has addressed the important content and process skills. She might offer additional lessons to cover any gaps in instruction after checking the rubric.

Rubrics also have the potential to enhance learning in several ways. When a teacher gives both the rubric and task at the same time, the rubric serves to direct student attention to the important concepts and skills that they should demonstrate in completing the task. However, teachers should not assume that students naturally understand how to use or interpret rubrics. It is likely that they will need guidance in learning how to use the rubric as an aid in structuring their initial responses and revising their work. Two approaches can help students use rubrics to improve their work.

First, having students construct a checklist can help them develop an understanding of evaluative criteria and how those criteria apply to their work. When initially assigning the project, the teacher might ask students: “When we are finished with this project, what will the really good ones look like? What things have we been learning that you might use in this project? What can you look for when you are done with your work to see if you have done well? Let’s create a list of what you would see in a good project.”

In this discussion the teacher is helping students develop a very sophisticated concept, the idea of evaluative criteria. Initially, students are likely to have difficulty coming up with descriptors (criteria). When students work to develop checklists, they begin to see what the teacher notices when grading papers. The students come to understand that pieces of work have certain characteristics that are connected to quality. They begin to develop an understanding of evaluative criteria.

When asked to develop such checklists for the first time, children may be preoccupied with effort as a criterion. They might say a product would be good, “if I tried really hard.” Other characteristics that students identify at first relate to surface features, such as neatness, and obvious ones, such as the right answer on a math paper. Teacher probing or suggestions will lead to an expansion of students’ understanding of important criteria and, over time, will lead students to develop insight into more sophisticated characteristics of good work.

Another beneficial classroom activity is having students use checklists to evaluate their own work or the work of peers. Once students have developed an understanding of evaluative criteria, they can begin to look at their work to identify instances where a specific skill is demonstrated. This is an important ability because it is key for students to be able to examine and reflect on their own work.

As students first begin to learn to apply checklists to their own work, it may be useful for teachers to use interviews or conferences to determine whether a child is able to use this new skill. For those students who are having difficulty in using checklists, teachers may use the conference to guide the student in applying the checklist to his work.

A special benefit comes about when students review the work of peers. As students examine the work of others, they come to understand, appreciate, and perhaps learn about strategies and approaches other students use in their work. The use of peer review, however, is psychologically and socially difficult for some learners. To promote this activity as a positive experience,

the teacher might propose the task as a treasure hunt in which the job is to look at a peer's work and find instances where the criteria have been met. Deficiencies might be noted as suggestions for revision.

In the same spirit, some teachers use "three pluses and a wish" in authors circles (Mills 1990), guiding students to find three strengths in a piece of work and recognize one way it can be improved. The use of this strategy ensures that students get positive feedback from peers as well as suggestions for improvement.

STEPS IN CONSTRUCTING RUBRICS AND CHECKLISTS

To assist you in creating scoring guides for use in your classroom, we offer the following steps. In this discussion a rubric will be used for grading the oral and written responses to the measurement task below.

Make a scale drawing of one of the rooms in your house. You are to select a scale that is appropriate for the size of the graph paper. Show important features of the room, such as doors and windows, using the symbols we learned about in class. Show the placement of furniture or appliances, also drawn to scale, and label them. Prepare to describe and explain your drawing to a small group of students next week. During your presentation, be sure to talk about the measurements of at least five features of the room.

Step 1. Decide what evaluative criteria are central to the task(s) to be graded.

The following is a list of tips that are useful for selecting evaluative criteria.

1. Make sure your expectations match curriculum standards.
2. Imagine what a good student response would look like.
3. Think about parts of the task students would find difficult.
4. Make sure that criteria are consistent with task directions.
5. Decide which task features will not be assessed.
6. Limit the number of criteria.
7. Decide whether the rubric will be specific or generic.

The first thing to consider are the curriculum standards or guides for the subject area. For the measurement task, the *Principles and Standards for School Mathematics* (NCTM 2000) would be useful to consult. The teacher should also consider what a good student paper and presentation would be like. The teacher might ask, "What do I expect to see in a good performance or product?" Based on prior experiences and current expectations, the teacher creates a mental image of what students will be able to do. At the same time the teacher should think about parts of the task that students

might find difficult. Such reflection would identify weaknesses in student responses and areas where further instruction would be beneficial.

The teacher should also consider the directions she has given to students as she develops evaluative criteria. For example, in the measurement task it would be unfair to use criteria related to the form or format of an answer—for example, meters and centimeters instead of feet and inches—unless such expectations were communicated to students in the task description. Sometimes, however, expectations are established as a classroom convention for all assignments, such as labeling all answers in mathematics, and need not be repeated in the directions. In short, a close match between the task description and the rubric is important.

The fifth tip suggests that teachers deliberately decide what will *not* be considered in the evaluation. Students can be overwhelmed with having too many criteria to consider as they complete tasks. Limiting the number of criteria allows them to focus on the really important learning outcomes. Teachers also benefit because their focus is on the most important areas of learning, so grading becomes less cumbersome.

Evaluative criteria that a teacher might use to grade student responses to the measurement task are shown in the following lists.

The Drawing

- accurate measurement with appropriate units
- appropriate use of a reasonable scale
- appropriate form showing title, scale key, labels, architectural notations
- neatness

The Presentation

- use of appropriate measurement terms
- accurate description of the measurements of five features in the drawing

Notice, criteria are defined for both the written product and the oral presentation. In developing this list, we reviewed NCTM Standards, thought about what good student responses would be like, and anticipated mistakes students might make. After writing the task and then thinking about student responses, it became clear that to assess students' work, the teacher would need to know what each item in the scale drawing represents. Thus, we added the requirement that pieces of furniture be labeled.

Also, the original task only had a written component. Students' verbal interpretation of the scale drawing also seemed to be an important dimension of the assessment. The scale drawing alone would not tell whether students could use appropriate measurement terms, thus we added an oral presentation to the task. We selected an oral presentation rather than a written explanation to reduce the reliance on students' writing skills in this assessment of mathematics knowledge and ability.

In developing any rubric, it is important to consider whether or not to assess surface features such as handwriting, spelling, grammar, and neat-

ness. In the case of scale drawings, we included neatness in the list of evaluative criteria. A decision may be made later to eliminate this criterion or to make it count for fewer points than other criteria.

Another matter to consider is whether the rubric is being designed to grade a single specific task (a task-specific rubric) or if it will be used repeatedly to grade different pieces of student work (a generic rubric). The decision about whether the rubric will be generic or task-specific influences the definition of evaluative criteria, a matter considered in Step 3. A task-specific rubric will be developed here; thus the focus will be only on measurement skills associated with a scale drawing of a room.

Step 2. Decide if the student work will be evaluated holistically or analytically.

As we mentioned earlier, the choice of analytic or holistic approaches is influenced by the intended purpose of the assessment. If the purpose is to provide an overall assessment that integrates the different elements of student performance, a holistic rubric might be preferred. If the teacher wanted to give students detailed feedback about their performance on each particular evaluative criterion, she would use an analytic scoring guide.

No matter whether you want to assess student work holistically or analytically, all the important criteria identified in Step 1 should be incorporated into the scoring guide. The difference is that the analytic approach would address each criterion separately, while a holistic approach requires the writing of a single description that integrates all of the criteria of student performance at each level.

An analytic rubric allows a student to perform poorly in one area and still receive an overall high grade. Since students earn points for several evaluative criteria, the loss of one point for an error in punctuation or capitalization will have a limited effect on the total score. In contrast, scoring with a holistic rubric may reduce the student's grade if the work has only one area that is weak, because all criteria are considered at once and any flaw has the potential to lower the overall judgment of the performance.

The rubric for the measurement task could be either holistic or analytic. We chose to use the analytic approach because there are a number of discrete skills to evaluate. For example, one student might use a reasonable scale but have inaccurate measurements, while another might select a scale that is too large but have correct measurements. In a holistic rubric both students might receive the same score, and the nature of the student difficulty would be lost.

If a holistic approach were appropriate, the next step would be to decide on the scale for the rubric, that is, how many levels it will include. This matter will be addressed in the discussion of Step 3.

The selection of an analytic approach to scoring rather than a holistic approach results in three possibilities for scoring guides. The checklist, point-allocated checklist, and analytic rubric are all analytic approaches. If

the teacher simply wants a record of whether or not the student demonstrated a skill, then a checklist is appropriate. In using a checklist, the teacher need only put blanks in front of each criterion on the list to provide a place for indicating whether or not the performance was acceptable.

If it is important to recognize different levels of performance related to some criteria, a point-allocated checklist might be better. In the case of the measurement task, some of the evaluative criteria lend themselves to assessing performance at different levels. For example, in doing a floor plan, students would have to measure the size of the room as well as several pieces of furniture. They would also have to decide if they should use centimeters or meters or inches or feet or some combination of appropriate units. What level of perfection is needed to receive a check? Must the work be flawless? Would one measurement error mean the student gets no credit? What if the room dimensions were given in inches rather than feet? Would the failure to use the traditional combination of feet and inches for the units deny credit even if all the measurements were correct? Clearly, awarding more than one point for this dimension of performance would allow variation in the scores of a student whose work is flawless and a student who places an eight-foot sofa along a five-foot wall but does everything else correctly.

A point-allocated checklist can also weight certain criteria to show that some areas of performance are valued more than others. For example, in grading essays, if five points are allocated to evaluating mechanics and grammar and twenty points are allocated for the content, the implicit message is that mechanics and grammar are important, but content is of greater importance in light of the learning outcomes being assessed by the task. The same strategy of weighting criteria can be used in an analytic rubric.

The difference between a point-allocated checklist and a rubric is that a rubric explicitly defines levels of performance for each criterion. Based on an earlier discussion, you can see that a point-allocated checklist for the measurement task might give three points for the criterion “accurate measurement with appropriate units.” As we discussed in the beginning of the chapter, an analytic rubric would explicitly communicate what is expected for a three-point response while a point-allocated checklist would not.

In determining which of the three types of scoring guides to use, the teacher must consider whether there is a need to define levels of performance explicitly. The scoring guide is finished if he is to use a checklist. For all other scoring guides, the next step involves decisions about selecting a scale for the guide.

Step 3. Develop rating scales for the scoring guide.

The *American Heritage Dictionary* offers one definition of *scale* that fits this discussion of rubrics: “A progressive classification, as of size, amount, importance, or rank.” The formation of a point-allocated checklist requires the teacher to decide the number of points—the amount—to give each criterion. The decision about how to allocate points is an actual communication

of the importance of each evaluative criterion. Thus, in grading a social studies report the teacher might allocate more points to criteria related to the content of the report and fewer points to form and format issues, such as grammar, mechanics, spelling, and neatness. But in an English class where students are writing historical fiction, more credit might be given to matters of writing style and less for the content.

In rubrics, numbers or categorical labels are used to designate the level of performance. Figure 4-4 used labels—Expert, Accomplished, Novice—rather than numbers to rank the performances demonstrated in students' limericks. It would have been just as appropriate to use numbers. In fact, the holistic rubric for mathematics communication in Figure 4-6 uses a combination of category names and numbers—Accomplished (3), Developing (2), Not Yet (1). Notice the larger the number, the more proficient the response.

The designation of a numerical scale that increases for higher levels of performance allows the teacher to combine the domains to get a single grade, as in an analytic rubric, or to use rubric scores from performance assessments in combination with scores from other types of assessment. We will discuss turning rubric scores into letter grades later in the chapter.

In selecting a scale, the teacher must determine how many levels of proficiency to define and decide what labels or numbers to associate with each level. The selection of evaluative criteria sometimes controls the number of levels in a scale because the teacher can only describe a limited number of performance levels for a particular criterion.

For example, the analytic rubric in Figure 4-3 has only two levels for scoring the number of lines criterion—the limerick either had five lines or it didn't. But it uses three levels for the evaluation of the other criteria, such as "contains descriptive words," because the teacher could define three levels of proficiency for them. Writing the descriptions that are to define the scale helps the teacher determine how many levels are appropriate for a given criterion.

In designing a rubric for the measurement task, the teacher would be able to define several levels for most of the evaluative criteria. Development of the descriptions usually begins with writing the narrative for the highest level of performance. The first step in writing the rubric in Figure 4-8 was to write a description for top performance on the criterion "Accurate measurement with appropriate units." In the development of this task-specific rubric, the evaluative criteria as well as the descriptors limit the use of the rubric to scale-drawing tasks. The descriptions of lower proficiency levels are based on what was written for the highest level. The lower levels are defined by how often students made measurement errors and their use of appropriate units. Using the language in the highest performance level to write the narrative for other levels helps the teacher create parallel structures for the descriptors for various levels of performance.

Four different descriptions could be written that clearly define unique levels of performance for most of the criteria. However, only three levels could be defined for the criteria "Neatness" and "Use of appropriate

Student Name:		Date:			
The Drawing	Level 1	Level 2	Level 3	Level 4	
<p>Accurate measurement with appropriate units</p> <p>Appropriate use of a reasonable scale</p> <p>Appropriate form showing title, scale key, labels, architectural notations</p> <p>Neatness</p>	<p>Many measurements have errors and inappropriate or missing units.</p> <p>Drawing is not done to scale.</p> <p>Many required elements of the drawing contain errors and some are missing.</p> <p>Lines, printing, and finished product are not done neatly.</p>	<p>Some measurements have errors and inappropriate or missing units.</p> <p>Only a few features in the drawing are correctly scaled.</p> <p>Some required elements of the drawing contain errors and some are missing.</p> <p>Lines, printing, or finished product are not done neatly.</p>	<p>Some measurements have errors or inappropriate or missing units.</p> <p>Most features in the drawing are correctly scaled.</p> <p>Some required elements of the drawing contain errors or some are missing.</p> <p>Lines, printing, and finished product are done neatly.</p>	<p>All measurements are correct and labeled with appropriate units.</p> <p>All features of the drawing are correctly scaled.</p> <p>All required elements of the drawing are present and correct.</p>	
The Presentation	Level 1	Level 2	Level 3	Level 4	
<p>Use of appropriate measurement terms</p> <p>Accurate description of the measurements of 5 features in the drawing</p>	<p>Terms are used incorrectly or units not mentioned.</p> <p>Descriptions of measurements are not given or are incorrect.</p>	<p>Some terms are used correctly.</p> <p>Measurements of 1 or 2 features in the drawing are accurately described.</p>	<p>All terms are used correctly.</p> <p>Measurements of 3 or 4 features in the drawing are accurately described.</p>	<p>Measurements of 5 or more features in the drawing are accurately described.</p>	

FIGURE 4-8 Analytic rubric for measurement task

measurement terms.” When the criteria in a rubric result in the use of a different number of levels for some criteria, those that have the higher number of levels are of course worth more points. These criteria, therefore, have a greater weight in determining the final grade. The teacher must consider if this effect on the final grade is appropriate. In our example, neatness will be only a three-point consideration in the grade on the scale drawing, while other criteria will count for four points.

If the number of levels in a rubric must be determined by the number of descriptions a teacher can write, it is important to consider what makes a good descriptor. Descriptors should:

1. address quantitative and/or qualitative aspects of the criterion
2. address the same elements of performance at each level
3. avoid generic adjectives such as *good*, *poor*, *weak*, and *excellent*
4. try to suggest needed improvements rather than noting deficiencies

First, the teacher must think about the quantitative and qualitative characteristics of performance that she will consider in grading the work. For example, in writing a persuasive letter, the author might be expected to give at least a certain number of reasons to support a point of view. The rubric might say, “At least three different reasons are given.” On the other hand, in a fictional narrative, a teacher may not want to count the number of adjectives in the piece at all. Thus, as in Figure 4–9, the descriptor might focus on qualitative characteristics: “Vivid descriptions are used effectively in some parts of the narrative.” In this example a quantitative consideration is also incorporated when the difference between the top two levels refers to frequency, using the phrases “throughout the narrative” in Level 4 and “in some parts of the narrative” in Level 3.

In mathematics a quantitative descriptor may focus on the number of computation errors to determine whether a paper is at one level or another. In other cases, the variation may be related to such qualitative elements as how organized, clear, or thorough the response is. A student’s response to a

Levels	Descriptors
4	Vivid descriptions of people, places, and emotions are used effectively throughout the narrative.
3	Vivid descriptions are used effectively in some parts of the narrative.
2	Attempts to use descriptions, but use of descriptive words is awkward or inappropriate.
1	Needs descriptive language.

FIGURE 4–9 Levels of performance for use of descriptive words

problem-solving task may be judged at a higher level because the explanation of strategies used is thorough, clearly illustrated, or well organized.

As mentioned earlier, use of parallel structures by addressing the same elements of performance in the descriptions at each level enhances the clarity of the language in a rubric. For example, the first criterion in the rubric for the measurement task (Figure 4–8) addresses both accuracy of measurements and use of unit labels at all four proficiency levels. At the lowest level, student work contains many measurement errors and unit labels are not appropriate, while at the highest level all measurements are correct and appropriately labeled. In a similar manner, the descriptor at each level of the first criterion addresses both accuracy of measurement and labeling.

In a holistic rubric several evaluative criteria are considered simultaneously and a change between proficiency levels may be triggered by variation in only one criterion. It is better to repeat the descriptions of the unchanged elements rather than to ignore them. The holistic scoring guide in Figure 4–10 illustrates this strategy. The descriptions of Levels 4, 5, and 6 include the statement “Lines, printing, and the finished product are done neatly.” Other elements of performance define the difference between an adequate and an outstanding response.

Also notice that the holistic rubric for the measurement task has six levels while the analytic rubric has only three or four levels for each criterion. More levels are needed in the holistic case because many different types of student performances must be considered in combination to select the score. Without a larger number of levels, it would be very difficult to decide what score a student should receive if he meets one performance expectation for a level for one criterion but does not meet another. Some teachers might use the same four-level schema for both the holistic and analytic rubrics and augment the scale by using plus and minus, similar to letter grading schema used in report cards, that is A+, A, A–, and so on).

When the description of a proficiency level involves more than one element, the use of *and* and *or* can distinguish between levels. Such is the case for the evaluative criterion “Tells a funny story” in the analytic rubric for scoring limericks (Figure 4–3). This criterion implied that the limerick both told a complete story and was funny. A limerick of high quality would be characterized by both elements. In defining the next lowest level of performance, the teacher might want to recognize that a limerick could be humorous but not tell a complete story, or it might tell a story that is not funny. Thus on a three-point rubric, the middle level of performance reads, “The story that the limerick tells is incomplete *or* is not funny.” The lowest level reads, “The story that the limerick tells is incomplete *and* is not funny.” Both the analytic and holistic rubrics for the measurement task make use of the conjunctions *and* and *or* to distinguish levels of proficiency.

The third characteristic of clear rubric descriptions emphasizes the importance of providing details in descriptors. The measurement rubric avoids the use of *good*, *weak*, and *poor* by providing specific descriptions of what the teacher expects to see in student work. The use of such words as *excellent* and *poor* fails to provide details that communicate what is expected to be

Student Name:		Date:
LEVELS	DESCRIPTORS	
Level 6	All measurements are correct and labeled with appropriate units. All features of the drawing are correctly scaled, and all required elements of the drawing are present and correct. Lines, printing, and the finished product are done neatly. In the presentation, all terms were used correctly and the student accurately described the measurements of 5 or more features in the drawing.	
Level 5	The student work is characterized by most of the following statements. There are some errors in measurements or inappropriate or missing units. All features of the drawing are correctly scaled, but some required elements contain errors and some are missing. Lines, printing, and the finished product are done neatly. In the presentation, all terms were used correctly and the student accurately described the measurements of only 3 or 4 features in the drawing.	
Level 4	The student work is characterized by most of the following statements. There are some errors in measurements or inappropriate or missing units. Most features in the drawing are correctly scaled, but some required elements contain errors and some are missing. Lines, printing, and the finished product are done neatly. In the presentation, most terms were used correctly and the student accurately described the measurements of only 3 or 4 features in the drawing.	
Level 3	The student work is characterized by most of the following statements. There are some errors in measurements and inappropriate or missing units. Most features in the drawing are correctly scaled, and some required elements contain errors while others are missing. Lines, printing, or finished product are not done neatly. In the presentation, only some terms were used correctly and the student accurately described the measurements of only 1 or 2 features in the drawing.	
Level 2	The student work is characterized by most of the following statements. There are some errors in measurements and inappropriate or missing units. Only a few features in the drawing are correctly scaled, and some required elements contain errors while others are missing. Lines, printing, or finished product are not done neatly. In the presentation, only some terms were used correctly and the student accurately described the measurements of only 1 or 2 features in the drawing.	
Level 1	There are many errors in measurements and inappropriate or missing units. The drawing is not done to scale, and many required elements contain errors and some are missing. Lines, printing, and finished product are not done neatly. In the presentation, terms were used incorrectly or units were not mentioned. The descriptions of measurements were not given or were incorrect.	

FIGURE 4-10 *Holistic rubric for measurement task*

achieved at a specific level. The labeling of a piece of work as *good* is subject to interpretation and confusion.

Why is *good* not good enough? Imagine substituting the descriptions *poor*, *good*, and *excellent* in the analytic rubric for limericks given earlier. The comparison of the original and new descriptions is provided in Figure 4–11. Students receiving a “Good” would not learn how to write a better piece. They may not understand that specific nouns, adjectives, and adverbs make the difference between an excellent and good performance. They may have seen their use of several adjectives as providing excellent descriptions and not realize that they also need to use adverbs or nouns to strengthen their limerick. The use of such general descriptors keeps the expectations for performance a secret.

A final consideration in writing descriptors is the use of language that is supportive of student learning by suggesting what the child needs to do rather than noting what is missing. For example, in Figure 4–4 the descriptors at the lowest level suggest “the limerick needs to contain specific nouns, adjectives, and adverbs” instead of saying “the limerick does not contain specific nouns, adjectives, and adverbs.” Sometimes attempts to use such language result in descriptions that seem unclear or unnecessarily awkward. The decision about language should be guided by consideration of which approach more clearly communicates expectations to students.

Contains descriptive words (Distinct language from from Figure 4–3)	1 No use of specific nouns, adjectives, and adverbs to paint a picture for the reader.	2 Attempt to select specific nouns, adjectives, and adverbs to paint a picture for the reader.	3 Effective selection of specific nouns, adjectives, and adverbs to paint a picture for the reader.
Or			
Contains descriptive words (Less effective substitution of generic language in descriptors)	1 Poor	2 Good	3 Excellent

FIGURE 4–11 Comparison of language in descriptors

The use of scoring guides often results in students asking, “What score do I need to get an A?” Families may also find rubric scores incomprehensible and want to know what they mean. In such cases, the teacher must be prepared to explain what rubric scores mean based on traditional indicators of student learning, such as letter grades.

Scores assigned with the rubrics and checklists discussed in this chapter can easily be associated with letter grades. Using the holistic rubric for limericks (Figure 4–4), the teacher may decide that a student earns a grade of C if the response is at the Novice level, a grade of B at the Accomplished level, and a grade of A at the Expert level.

When using an analytic rubric, the teacher can determine a grade by adding the scores for the evaluative criteria and then designating a letter grade for a certain span of scores. For example, in the analytic rubric for the measurement task (Figure 4–8), the highest possible total score would be 22—four criteria with a high score of 4 and two criteria with a high score of 3. A teacher might decide that scores of 19 to 22 will be A’s, scores of 16 to 18 will be B’s, and so on. There is no single right way to make such conversions. It’s just important that the teacher communicates to students and their families the meaning of scores in terms of the expected standards of performance and grades.

SPECIAL CONSIDERATIONS

The design of a scoring guide should take into account differences in learners. The ages of students and their language fluency, for example, are considerations. When scoring guides are used with pre- or emergent readers, teachers sometimes use pictures instead of words or numbers to represent the concepts to be learned. To introduce children to scoring guides, it is probably best to use checklists at first and then move toward rubrics, because the use of scales and weights in creating scoring guides could be confusing to young learners.

For example, a checklist for four- or five-year-olds who are learning about directions such as right, left, up, and down might use an icon approach as shown in Figure 4–12. Notice that the use of words and icons will help the child and family communicate about what has been learned.

Other icon-based checklists could be used to show skills such as using scissors (✂), knowing a phone number (☎) or address (📮), and using a computer keyboard (⌨) or the mouse (🖱). Imaginative use of clip art and computer fonts makes it possible for teachers to create icon-based scoring guides for young learners. The teacher will, of course, have to explain the significance of the symbols and what the check means when first using such checklists.

Name: Jacob C.	
Date: April 12	
←	(left)
→	(right)
✓ ↑	(up)
✓ ↓	(down)

FIGURE 4-12 *Icon-based checklist*

Once students become familiar with checklists, teachers may want to begin using an icon-based rubric that will communicate variation in levels of performance. Figure 4-13 uses multiple stars to indicate a greater level of proficiency with counting. The specific skills relate to knowing counting words, knowing symbols, and being able to count collections of objects. Notice that the use of both symbols and words results in a scoring guide that can be interpreted by both children and adults.

POWER OF RUBRICS AND CHECKLISTS

The types of scoring guides described in this chapter do much to enhance the teaching and learning environment. As we mentioned earlier, teachers have an easier time evaluating student work and being fair in their evaluation when using a well-constructed scoring guide. At the same time, the use of rubrics and checklists supports student learning. Students develop a clear understanding of what is expected if they receive both the task directions and the evaluation rubric at the same time. The rubric serves to direct stu-

NAME: Avery C.		Date: Oct. 15, 2000	
Knows counting words	★	★★	★★★
	“one” to “five”	“one” to “ten”	“one” to “fifteen”
Knows symbols	★	★★	★★★
	1, 2, 3, 4, 5	1, 2, 3, ..., 10	1, 2, 3, ..., 15
Counts objects	★	★★	★★★
	up to 5	up to 10	up to 15

FIGURE 4-13 *Icon-based rubric*

dent attention to the important concepts and skills they should demonstrate in completing the task. Such open communication with students can serve to reduce anxiety. Further, the teacher-student relationship during assessment is changed from an adversarial one to a pedagogical one. The assessment is not an occasion for the teacher to catch the student in an error, but an opportunity for students to demonstrate what they have learned.