MASTERING ASSESSMENT: A SELF-SERVICE SYSTEM FOR EDUCATORS

The Role of Rubrics in Testing *and* Teaching

W. James Popham

Routledge
Taylor & Francis Group
New York London

PREFACE

This booklet, one in a series of 15 booklets contained in *Mastering Assessment: A Self-Service System for Educators (MA)*, was written chiefly for those who currently teach in our schools. Because I was once a public school teacher, I have lasting respect for our nation's teachers. Thus, I have not tried to write a series of tiny *textbooks* wherein an author attempts to teach a reader. In this booklet, therefore, you will find no practice exercises, no end-of-chapter quizzes, and no interminable list of additional references. Instead, I tried to write each booklet in the form of a colleague-to-colleague conversation—a somewhat one-sided conversation to be sure. What I attempted to do in every booklet was explain some assessment ideas I regard as important for today's teachers to understand. I attempted to provide those explanations employing the language I'd use if I were sitting with a fellow teacher in a faculty lounge. Even so, you'll occasionally encounter a dollop or two of irreverent whimsy in these *MA* booklets because I seem altogether incapable of controlling my silliness propensities—in faculty lounges or elsewhere. Given my hope that readers will acquire several assessment-related *understandings* from each *MA* booklet, in the spirit of collegial candor I have laid out those anticipated understandings at the start of every booklet. Moreover, at each booklet's conclusion I have briefly reiterated what I regard as the essence of those understandings. After all, even colleagues sometimes need to repeat themselves.

WJP

Anticipated Understandings

After reading this booklet, you should understand:

- ***What a rubric is and why its use can be beneficial from both an assessment and an instructional perspective.***
- ***What the three essential elements of a rubric are.***
- ***How to distinguish between instructionally worthwhile and instructionally worthless rubrics.***

Rubricity is rampant in the land. Putting that same point another way—and using only real words—these days you'll find enormous numbers of educators relying on *rubrics*. Indeed, it is a rare occurrence for a full week to go by without a teacher's hearing a colleague or a student refer to a rubric. But, unlike many educational fads that seem to surround our schools for a few years and then scurry into obscurity, rubrics are definitely here to stay. That's why you need to learn about them. More specifically, that's why you need to learn what rubrics are and how they can help you do a better job of testing your students *and a better job of teaching your students*.

WHAT'S IN A NAME?

Let's get immediately into what a rubric is and the source of its fairly cryptic name—a name often confused with the label of a popular puzzle (*"Rubik's Cube"*). A rubric is a *scoring guide*. It is a scoring guide used to evaluate the quality of students' work. Currently, the most widespread use of rubrics can be seen in the evaluation of students' written compositions. Indeed, the earliest applications of rubrics in the United States dealt almost exclusively with the scoring of written compositions.

Rubric: A scoring guide used to judge students' work

To give you a general idea of what those early rubrics looked like, in Figure 1 I've presented the chief features of a rubric for judging students' compositions that I ran across in the 1970s. Interestingly, today's rubrics evaluate students' written compositions along much the same lines.

Figure 1 lists five evaluative factors for scorers to use in judging a composition's quality. Although the earliest rubrics were employed to score large numbers of student compositions in districtwide or statewide assessments, it should be apparent that a classroom teacher could use such a rubric for judging the written compositions of the teacher's own students.

Figure 1. An Early Rubric for Judging Students' Written Compositions

Directions to Scorers: Assign a *strong, acceptable,* or *weak* to a student's composition for each of the scoring factors below, then arrive at an overall evaluation of the composition based on your per-factor judgments.

- ***Content:*** Are the information and ideas in the composition appropriate?
- ***Organization:*** Is the composition's structure suitable for the writer's purpose?
- ***Word Choice :*** Is the vocabulary consonant with the writer's purpose?
- ***Fluency:*** Are the composition's sentences varied and powerful?
- ***Mechanics:*** Are mistakes in spelling, capitalization, punctuation, or grammar likely to distract a reader of the composition?

In addition to written work, rubrics can be used to evaluate the quality of any student-generated *product* or any student-generated *behavior*. For instance, when a student in a woodworking class creates a pair of mahogany bookends (*a product*), the quality of those bookends can be appraised more accurately when the instructor relies on a rubric to make a systematic judgment rather than trying to judge the bookends using a more haphazard approach.

Student Product: Something material a student creates to demonstrate the student's learning

Student Behavior: A student's performance intended to demonstrate what the student has learned

Similarly, when teachers in a speech class must evaluate their students' "demonstration speeches" (*a behavior*), those teachers will do a more defensible evaluative job when they employ rubrics. Thus, whether it is the evaluation of students' products or behaviors, teachers who rely on rubrics almost always do a better evaluative job than teachers who don't.

One useful way of thinking about the applicability of rubrics to the evaluation of students' work is to distinguish between test items that call for students to *select* their responses versus test items that call for students to *construct* their

responses. These two types of items are referred to, not surprisingly, as *selected-response* items and *constructed-response* items. The most common varieties of selected-response items are the sorts of multiple-choice and true-false items that teachers have been using in their tests ever since the introduction of papyrus. The most common kinds of constructed-response items, of course, are short-answer and essay items. When Plato asked Aristotle to "Define life, orally, in 100 words or less," that test item also fell into the constructed-response category.

See also *Selected-Response Tests: Building and Bettering*

Selected-Response Items:
Test items requiring students to choose from two or more options

See also *Constructed-Response Tests: Building and Bettering*

Constructed-Response Items:
Test items requiring students to generate their answers

There's clearly little need for rubrics when teachers are scoring selected-response items. After all, if choice C is the correct choice, and a teacher possesses even rudimentary alphabetical skills, then scoring a student's response to an item isn't all that challenging. Rubric-based scoring applies chiefly to the way students complete constructed-response tasks (that is, items), but a teacher needs to be judicious in deciding which constructed-response items warrant their use. For instance, it would be silly to whip up a full-blown rubric to evaluate students' responses to the fourth-grade, constructed-response social studies item below:

What are the names of the two major citizen-elected federal lawmaking bodies of the United States government?

(1) _____

(2) _____

WHEN TO USE A RUBRIC?

Let me preview you a bit on an important insight I hope you'll gain by reading this booklet: the creation of a really good rubric requires substantial effort. Mediocre rubrics, unfortunately, can be generated by teachers rapidly—often without any cerebral

Rubric is a term that has a lofty etymological lineage. It comes to us all the way from the Middle Ages. In those days, an era clearly prior to the arrival of the printing press—multiple copies of books were created almost exclusively by Christian monks who, laboring in their monasteries, copied scriptural literature from dawn to dusk—in Latin, of course. Well, the beginning of each major new section of these monk-copied books usually was written in large red letters. That distinctive color was supposed to alert the reader to the book's upcoming new section. Because the Latin modifier for red materials is *rubrica*, the word rubric soon began being used as the label for a section of a book or, by extension, simply as a category—in books or elsewhere. That's how the term *rubric* came to us. There are even derivative forms of the word that you might enjoy tossing around while waiting for a faculty meeting to start. For example, *rubricate* is a verb describing what a person does who adds red letters to a manuscript. Come to think of it, when I was a teacher who graded my students' essays with zealous flourishes of my red Number 2 pencil, I apparently was an in-the-closet *rubricator.*

activity whatsoever. I have personally seen many rubrics that were apparently developed by teachers who must have had other things on their mind. But a first-rate rubric—one that can help a teacher score students' work and one that can also help a teacher do a better instructional job—requires some serious thinking.

I recommend—no, actually, I plead—that you not develop rubrics for use with any old constructed-response test items that you use in your classes. Instead, restrict your use of rubrics to assessing students' mastery of only the *most important cognitive skills* you want your students to acquire. It is, based on my battle-battered experience, far better for teachers to rely on a handful of really fine rubrics than to employ dozens of decidedly dismal ones.

Let me give you an example or two of the sorts of cognitive skills I regard as warranting the use of rubrics. Let's start

with a biology teacher who wants to get students to understand "the inquiry process." To see if her students can employ such a process as they design and carry out original investigations, she might devise a fairly elaborate, multi-week *performance test*. To help the teacher evaluate her students' mastery of that obviously high-level science skill, a rubric should surely be used.

Performance Test: An assessment task typically requiring students to create a somewhat elaborate constructed response

As a second example, think of a middle-school history teacher who wants his students to be able to "use history's lessons to predict today's events." The teacher gives his students a series of current-day problem-situations facing society, then asks the students to predict what will happen in that problem-situation and defend their predictions in essays predicated on historical parallels. To judge the caliber of the students' history-soaked essays, a rubric would really come in handy.

You'll note, earlier, that I encouraged you to use rubrics for the appraisal of students' *cognitive skills*, not their *knowledge*. Knowledge, using the 50-year-old definition supplied to us by Benjamin Bloom and his colleagues (1956), refers to memorized information such as students' ability to recall facts, dates, terminology, and so on. Well, I definitely don't want to discount the importance of students' acquiring all sorts of nifty knowledge. In fact, without students' mastery of such knowledge, it is often impossible to promote students' acquisition of high-level cognitive skills. Knowledge is, unarguably, a worthwhile thing for students to possess.

Cognitive Skill: A specific intellectual competency that individuals can apply in appropriate situations

Knowledge: Memorized information, for example, recollectable facts, dates, principles, and terminology

But students' knowledge is almost always more efficiently assessed using selected-response items or, perhaps, the kinds of constructed-response items (such as short-answer ones) that really don't require the use of sophisticated rubrics.

Moreover, teachers shouldn't focus their instructional energies on promoting the mastery of so many cognitive skills that they—and their students—become almost literally overwhelmed. As noted in the preface of this booklet, I've tried to write it using a colleague-to-colleague style, and I understand that reasonable colleagues can differ on the issue of how many instructional targets a teacher can *successfully* address. So please don't regard my point of view as anything more than what it is—a point of view.

Nonetheless, I definitely lean toward lean. You may not. You see, I'd rather have a social studies teacher attempt to get students to master a half-dozen terrifically important cognitive skills—over the course of a semester or a school year—than attempt to get students to master 20–30 somewhat important cognitive skills. I believe, based on about five decades of interacting with in-the-trenches teachers, that "less does, in fact, turn out to be more." I want students to master a small collection of super-significant skills—really master the skills—so that those students can subsequently employ such super-significant skills not only during school, but also in their post-school lives.

WHY USE A RUBRIC?

All right, I've indicated the sorts of skills for which teachers ought to use rubrics—at least the sorts of skills that seem to make sense for busy teachers to use. (Incidentally, if you are reading this booklet and you're one of the few teachers on earth who is *not* busy, keep it a secret.) These days, more than ever before, it is important for teachers to have an appropriate answer to the following fundamental question: "Why use rubrics at all?"

Well, one obvious *incorrect* answer to that query is most certainly, "Because they're popular." Nor would you receive all that much credit if your answer was, "To score students' work." You see, the latter response yields only partial-credit because, in addition to serving as a guide for scoring, a well-formed rubric can be a potent tool to help you teach more effectively. So, a praiseworthy answer to the question of "Why use rubrics at all?" is that their use helps a teacher do a better job of scoring students' work *and* a better job of teaching those students.

I've been messing around with education for a long time. During most of my career, whenever I was asked to identify the most important factor in making a teacher's instruction effective, I would always answer "giving students lots of engaged time-on-task," that is, the number of opportunities teachers give students to practice the skills those students are supposed to master. There is all sorts of empirical evidence to back up that view.

Recently, however, I've become convinced that the single, most important factor in making a teacher's instruction effective is *clarity of curricular intent*. Teachers who truly understand the nature of the curricular outcomes they want their students to master will, almost certainly, be able to do a better instructional job than will teachers who have only a fuzzy notion of what it is they want their students to achieve. And, happily, that's where a well-formed rubric comes in. A

properly constructed rubric can remove the fuzz from just about any peach in your curriculum cart. That's why rubrics aren't, fad-like, soon going to fade from the scene. A properly constructed rubric can help teachers better understand the nature of the often complex cognitive skills their students are supposed to attain.

If teachers have a better grasp of their curricular targets, they will almost certainly design lessons that give students (1) more on-target practice and (2) support in the acquisition of any precursive subskills or knowledge those students will need for ultimate skill-mastery. Moreover, teachers who are clearheaded about where they are headed curricularly will surely be able to supply more accurate explanations to their students during an instructional sequence. Have you ever tried to explain what a gerund is to students when you don't understand the difference between a gerund and a gerbil? As a first-year teacher, I once tried to teach my students about gerunds without genuinely understanding what made a gerund jiggle. It was a staggering instructional disaster.

A colleague recently told me that, "The primary purpose of a rubric is to help students learn to pay attention to what is important to pay attention to." I concur.

And, while we are on the topic of how rubrics can help make instruction more successful, there is no reason that students should be shut out of this clarification game. Almost all well-formed rubrics can be rewritten in age-appropriate language so that students themselves can become more clear-headed about the nature of the skills they're supposed to be mastering. For older students, rewriting a rubric is often not even necessary. Obviously, clarity of curricular intent can help teachers, but such clarity can also be a boon to students.

Some teachers mistakenly believe that the use of rubrics somehow will allow them to dodge the tricky task of giving grades to students. Rubrics do all sorts of good things for educators, but they can't (1) cure the common cold or (2) relieve teachers of the chore of deciding which students get which grades. Sorry!

A rubric that's put together properly will permit educators to arrive at more accurate inferences about students' status regarding mastery of the skills being assessed. But there's also an enormous *instructional* yield that can be derived from a properly constructed rubric. This is why rubrics can be useful for testing *and* teaching. That's the reason you need to understand how to whomp up a well-formed rubric. I'll now turn to

that issue by describing the three essential features of a properly whomped-up rubric.

WHAT'S IN A RUBRIC?

See also *How Testing Can Help Teaching*

There are three features in a well-formed rubric.

- ***Evaluative criteria:*** These are the factors to be used when judging the quality of a student's response.
- ***Quality distinctions:*** For each evaluative criterion, different levels of quality in a student's response must be described.
- ***Application strategy:*** Users of the rubric are told whether a student's response is to be judged using the evaluative criteria collectively or on a criterion-by-criterion basis.

Evaluative Criteria

Quality-determining factors. To determine the quality of a student's constructed response, for example, to judge the excellence of a student's participation in a formal debate, we need to decide on the factors to use when arriving at a defensible qualitative judgment. If, for instance, a debate coach has concluded that a debater's statements should be evaluated, at least in part, based on their clear relevance to the proposition being debated, then "proposition-relevance" would be an appropriate evaluative criterion to include.

Evaluative Criteria: Factors used when using a rubric to judge a student's response

Put another way, evaluative criteria are the *things* we use when judging the merit of a student's response. If there are five important things that a teacher believes should be used to evaluate students' essays, then the rubric that the teacher builds to score those essays should contain five evaluative criteria reflecting those five things.

How many evaluative criteria? There's another potential problem lying in wait when a rubric-builder decides on evaluative criteria—and that problem is linked to the optimal *number* of the evaluative criteria to use in a rubric. As you can probably guess from my previous applause for a "less is more" isolation of the cognitive skills to be taught and tested, I will once again lean toward lean. Remember, a rubric's evaluative criteria need to be thoroughly mastered by students so that, when a teacher's instruction has been finished, the students will be able to carry

away a solid evaluative framework from the class—a framework to help them evaluate their own subsequent use of that skill. In my experience, the number of evaluative criteria for a rubric ought to be somewhere in the neighborhood of three to six. I recommend selecting a modest number of the most important factors when judging a student's skill-mastery—doing so will make the rubric more manageable—from both an instructional and an evaluative perspective. If a rubric contains a dozen or more evaluative criteria, then students—and teachers—tend to get lost. If a rubric contains only one evaluative criterion, there are usually some significant evaluative factors that have been omitted.

The rubric-builder needs to identify what key factors should be involved when judging the worth of a student's work. But many novice rubric-builders, as they become more and more conversant with the skill being assessed, will see more and more subtle factors by which to judge students' responses. Familiarity, in this instance, breeds not contempt but, instead, may entice the rubric-builder to incorporate too many evaluative criteria. These rubrics invariably collapse because they are overloaded with too many things.

Concise labels. Finally, because these evaluative criteria will typically be the focus of the teacher's instruction—and the backbone of a student's emerging skill—mastery, try to

give each evaluative criterion a brief but descriptive title. For example, if two of the evaluative criteria being used in a rubric for scoring students' written compositions are "organization" and "mechanics," those two labels convey a reasonable picture of what the focus of each evaluative criterion is.

To illustrate, Indiana University's Roger Farr, one of the nation's leading language arts authorities, recommends that

rubrics used to evaluate students' "purposeful reading" skills should contain the following three evaluative criteria (Farr, personal communication, 2003):

- ***Accuracy:*** How *accurate* is the reader's grasp and use of the text?
- ***Relevance:*** How *relevant* is the textual information used by the reader to fulfill the reader's purpose?
- ***Sufficiency:*** Does the reader demonstrate and use a *sufficient* amount of the text to fulfill the task?

Quality Distinctions for the Evaluative Criteria

Evaluative criteria are next to worthless unless a rubric lets the rubric-user know—for each evaluative criterion—what's good and what's bad. By setting forth *quality distinctions* for each evaluative criterion, we let the teacher know what to look for when evaluating students' responses. Assuming students have access to the rubric in some form, we also let students know what's needed if they are to come up with high-quality responses.

Quality Distinctions: Descriptions of different qualitative levels for each of a rubric's evaluation criteria

The fewer words that you can use to spell out quality distinctions for each of a rubric's evaluative criteria, the better. The two most common ways of supplying qualitative differentiations are:

- ***Two-Directional Quality Definitions:*** Describing only the general nature of highest-quality responses and lowest-quality responses
- ***Numerical Gradations:*** Providing quantitative score-point levels *or* performance-level descriptors for each evaluative criterion

The "two-directional quality definitions" approach simply sketches the essence of what should be regarded as students' best and worst responses to a particular task (or test item). For instance, if one of the evaluative criteria in a rubric for evaluating students' oral, in-class presentations is "eye-contact," the two-directional method of explicating qualitative differences for that criterion might be something along the following lines:

An Illustrative Two-Directional Quality Definition

☞ ***Evaluation Criterion:*** **Eye-Contact.** During a student's oral presentation, this criterion is best satisfied when students maintain almost constant eye-contact with most, if not all, members of the class. This criterion is least well satisfied when the speaker rarely, if ever, looks class members directly in the eye.

In contrast, a more quantitative approach spells out per-criterion quality distinctions based on "numerical gradations." Suppose, for example, that such a scheme had been used in order to characterize quality-levels in students' essays based on the number of errors made in "mechanics." Let's assume that the evaluative criterion of "mechanics" has been defined as "spelling, punctuation, grammar, and word usage." Therefore, an example of a numerical-gradation approach to defining the qualitative differences of students' responses for such a criterion might look something like this:

An Illustrative Numerical Gradation Approach

Evaluative Criterion: **Mechanics.** *The following point-allocations are to be awarded according to the number of clear-cut errors found in a student's essay:*

4 points	*Zero to two errors*
3 points	*Three to six errors*
2 points	*Seven to ten errors*
1 point	*Eleven to fifteen errors*
0 points	*More than fifteen errors*

This kind of numerical gradation, of course, could easily be converted into performance-level descriptors. For instance, instead of allocating students zero-to-four points on each criterion, the student might be assigned performance-level classifications such as "strong," "average," and "weak," or even such popular descriptors as "advanced," "proficient," "basic," and "below basic."

In choosing between these two methods of explicating the quality distinctions for a rubric's evaluative criteria, there is no one preferred approach. Indeed, those who are experienced in rubric-construction typically seem to make their choices on a skill-specific basis, that is, according to the skill being assessed as well as the nature of the particular evaluative criteria that have been chosen for use in the rubric.

Two real-world illustrations. Presented in Figures 2 and 3 are illustrative rubrics developed by district-level educators. Both briefly describe the skill being assessed, the evalua-

tive criteria, and the way of determining quality for each evaluative criterion. For the summarizing skill in Figure 2 there are four evaluative criteria and the quality distinctions for those criteria are set forth by identifying the nature of higher-quality and lower-quality responses for each evaluative criterion. In other words, a two-directional approach to quality distinction has been adopted in this example.

The two-directional strategy incorporated in the Figure 2 rubric, of course, may be less than completely precise in guiding teachers when they score students' summaries or, for that matter, in helping students know *for sure* whether they have personally written good summaries. However, even though not totally definitive, for both assessment and instructional purposes, these sorts of two-directional quality definitions are typically quite helpful.

Figure 2. An Illustrative Rubric for Judging Students' Summarization Skill

Cognitive Skill: Students will be given a multi-page expository reading selection, asked to read it, then generate a succinct and accurate summary of the selection.

Evaluative Criteria and Quality Distinction:

1. *Recounting of Important Elements:* The student's summary either reiterates or paraphrases the reading selection's essence and highlights significant points. Ideally, the summary would be written in the student's own words.

The highest quality response will recount the most important points in the selection and include few unimportant points. The quality of responses will decrease insofar as they (a) fail to include the most important points or (b) include many unimportant points in the selection.

2. *Brevity:* The reading selection's important content is concisely recounted.

The highest quality response will be the summary that most briefly recounts all of the reading selection's important points. Lengthier summaries will be considered of lower quality.

3. *Organization:* The summary is organized in a manner appropriate to the reading selection's content, for example, chronological, logical, or order-of-importance.

High quality summaries will be well organized; lower quality summaries will be less well organized.

4. *Mechanics:* The summary contains acceptable spelling, punctuation, and grammar.

High quality summaries will contain no errors in mechanics that diminish the effectiveness with which the summary communicates to the reader. In lower quality summaries, mechanical errors will distract the reader.

In Figure 3 you will find a rubric intended to evaluate a student's skill in identifying a theme in a narrative passage. Note that, in this instance, the student is supposed to read a narrative selection containing one or more themes, then satisfactorily describe the nature of a theme in the assigned selection. In this rubric there are two evaluative criteria used, namely, "accuracy" and "appropriate level of generality." Both of those evaluative criteria have been *combined* in order to generate four performance levels. This rubric illustrates just one of the ways that educators have attempted to delineate quality-levels for a rubric's evaluative criteria. Please recognize that this rubric's designer could have used point-allocations rather than performance-level classifications.

Figure 3. An Illustrative Rubric for Appraising Students' Theme-Identification Skill

Cognitive Skill: Students will be provided with a multiparagraph passage of 400–1000 words. Students will be asked to read the passage, then identify its theme (if the passage has only one theme) or *a* theme (if the passage has more than one theme).

Evaluative Criteria:

1. *Accuracy:* The degree to which the student's statement of a passage's theme coincides with one of the passage's actual themes.

2. *Appropriate Level of Generality:* The degree to which the student's statement of a passage's theme is sufficiently general so that it is applicable in other settings.

Performance Standards:

Advanced: The student's statement of the theme is completely accurate and is stated as generally as is warranted by the content of the narrative passage.

Proficient: The student's statement of the theme is completely accurate, yet stated at a level of generality that is somewhat too broad or too narrow *or* is almost completely accurate, yet stated at an appropriate level of generality.

Basic: The student's statement of the theme is either accurate, yet too narrow/broad *or* inaccurate, yet represents an appropriate level of generality.

Below Basic: The student's statement of the theme is both inaccurate and given at an inappropriate level of generality.

In both of these illustrative rubrics, a clear effort has been made to keep the reading level relatively modest. It would be possible, of course, to spell out a series of more quantitatively constraining performance levels for each evaluative criterion, but frankly, the complexity of such elaborate rubrics is usually off-putting to most teachers.

I find myself personally drawn to the sort of two-directional quality definitions seen in the Figure 2 rubric that's focused on summarizing. This kind of directional guidance seems, in my experience, to be sufficient for most teachers and students. Incidentally, teaching students to be able to produce concise and accurate summaries of what they have read is definitely a nontrivial teaching challenge.

Application Strategy

Let's suppose you have developed a really spectacular rubric that's suitable for assessing the degree to which a student has mastered a high-level cognitive skill. Your spectacular rubric, of course, contains a small number of significant evaluative criteria (let's say four or five) and, along with each evaluative criterion, you've supplied guidance to the rubric-user about how to apply those criteria when deciding on the quality of a student's work. Hats off to you—so far. But you still must make a decision about how those evaluative criteria and quality distinctions should be applied. And this is why the rubric's *application* strategy needs to be determined.

Application Strategy: Use of a rubric's evaluative criteria holistically (all together) or analytically (one-by-one)

Two basic application strategies. You can either apply a rubric *holistically* or you can apply it *analytically*. When a rubric is used holistically, the rubric-user tries to take into consideration *all* of the evaluative criteria and their accompanying

qualitative distinctions, but then makes one overall, that is, one *holistic*, judgment. This particular application of rubrics might stem, at least in part, from some variant of *holism* theory—the philosophic view that whole entities, because they are themselves fundamental components of reality, possess an existence beyond the mere sum of their parts. However, although that sort of high-toned rationale sounds really ritzy, there's a far more simple reason for using rubrics holistically: holistic scoring saves tons of time.

To employ a rubric holistically, a test-scorer tries to become really familiar with the rubric's evaluative criteria and the way those criteria are to be applied qualitatively. Then the scorer judges the student's response (for example, the write-up of a student's personally designed science experiment) and, while trying to keep all of the rubric's evaluative criteria in mind, comes up with a single, *total* judgment regarding the student's response. Holistic scoring of students' responses can take place quite rapidly, yet with considerable consistency.

Holistic Scoring: When a rubric's evaluative criteria contribute to a single, overall judgment of quality

Analytic Scoring: When a rubric's evaluative criteria are applied separately in judging students' work

In contrast, when a rubric is to be used *analytically*, the scorer must render a separate judgment for *each* of the rubric's evaluative criteria. To illustrate, if a rubric contains three evaluative criteria, then a scorer would make three separate per-criterion judgments when appraising a student's response. Sometimes these per-criterion evaluations are then amalgamated into a single, overall evaluation using rubric-specified step-by-step procedures. Often, however, there is no attempt to aggregate the separate evaluations.

These per-criterion evaluations can be rendered in numerical form (for instance, from 1 to 5) or as performance levels (for instance, "proficient," etc.), but I hope you recognize that a rubric-user needs to make a series of *separate* judgments for each of a rubric's evaluative criteria. This clearly requires more scoring-time than would be needed to supply a solo holistic judgment.

Advantages and disadvantages. The advantage of analytic rubrics, of course, is that they supply diagnostic data of considerable utility. Let's say you're a teacher who discovers that your students failed to master a significant skill chiefly because they tumbled terribly on *one* of a rubric's five evaluative criteria. (You see, you would definitely know this *if* the rubric had been applied analytically.) Then you and your students could get cracking on addressing students' conversance with

the evaluative criterion that led to their weak performances. You and the students would know precisely what needs to be addressed instructionally. However, had the rubric been applied holistically, you'd never know where it was that your students (and your instruction) had gone awry.

And there, in a walnut shell, are the two factors you need to consider when deciding whether to opt for an analytic or holistic application of a rubric. Holistic scoring takes less time to do, but provides no diagnostic data. Analytic scoring takes more time to do, but can supply both teachers and students with instructionally useful diagnostic data.

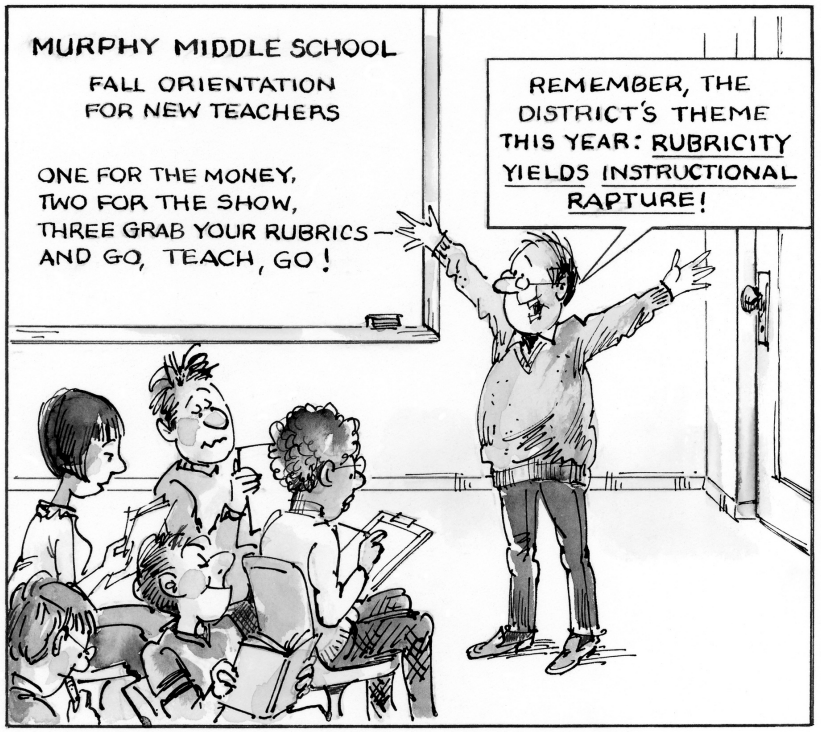
This third attribute of a well-formed rubric, that is, its application strategy, should make it apparent that rubrics can often be used in *either* way. It's true, however, that if the rubric-builder believes the rubric is likely to be used analytically, it is especially important to spell out even more lucidly what each evaluative criterion's quality distinctions are.

AN UNWARRANTED REVERENCE FOR RUBRICS

When most of the early rubrics appeared in our school a couple of decades ago, they focused on the evaluation of students' written compositions, and with few exceptions they were quite good. I suspect that these early composition rubrics were pioneered by savvy English teachers who had personally scored so many compositions through the years that they really understood the essentials of what was needed to accurately and consistently judge a student's written work.

> ***I am convinced that, because of more that 20 years' worth of high-stakes composition tests accompanied by first-rate rubrics, U.S. students are more skilled writers than they were when I was a high-school teacher and tested my students' writing prowess by seeing if they could diagram and punctuate sentences.***

Many modern rubrics, however—whether constructed by classroom teachers, by state curriculum specialists, or by assessment specialists from test-development agencies—are dreadful. They don't help teachers score students' work with any sort of objectivity or consistency. Beyond that, many of those rubrics are instructionally feckless. Teachers can't pick up any defensible ideas from these weak rubrics about how to design or deliver instruction for the skill allegedly assessed via the rubric.



I'd estimate that more than three-fourths of the rubrics I've encountered in the last 10 years should have been sent directly to a shredder instead of being used to evaluate students' work or to help teachers think about their instructional plans. Putting it simply, rubrics can be good and rubrics can be bad. With this simpleminded but significant insight in mind, I now want to direct your attention to two really rotten ways to construct rubrics and to one quite elegant way to do so.

RUBRICS—THE RANCID AND THE RAPTUROUS

Three kinds of rubrics currently roam our educational landscape. The first two, while seemingly suitable at first glance, are either scoring disasters or instructional nightmares. The final kind of rubric, the one that I'll be touting, works wonderfully—especially in its contribution to getting students to master really high-level cognitive skills. I want to start out by digging into an approach to rubric-building that initially seems sublime but soon sinks into a cesspool of generality.

Hypergeneral Rubrics

The main mission of rubrics, you will recall, is to clarify. If a rubric fails to clarify how to judge the worth of a student's response to a task, that rubric fails to satisfy its *raison d'être*. Sadly, a fair number of rubrics rely on per-criterion quality distinctions that are too general to do anyone—teachers or students—any genuine good. I refer to such excessively inexact scoring guides as *hypergeneral rubrics*. And that's precisely what they are: too blinking general.

Hypergeneral Rubrics: Excessively inexact, often vague, scoring guides

A vague rubric, though doubtlessly well-intentioned, is of little value to anyone. All that students and teachers really know from them is that a good response is, well, good—and that a bad response is, well, the opposite. That sort of general guidance regarding quality is far too gunky to be of any genuine assistance to teachers or students.

Here's an example of an illustrative excerpt from a hypergeneral rubric dealing with students' mathematical problem-solving skills. The student is given a problem-solving task, then asked to generate an appropriate and original solution. Notice that the evaluative criterion involved in this illustration is "Application of Relevant Tools" and that only three levels of quality have been set forth in the rubric. Note, in particular, the specific way that the rubric's developers have tried to distinguish among the three levels of quality when judging students' responses according to this evaluative criterion.

An Illustrative Excerpt from a Hypergeneral Rubric

Evaluative Criterion: Application of Relevant Tools.

When judging students' problem-solutions, the following three levels of quality should be employed to evaluate the student's application of relevant problem-solution tools:

- ***Distinguished:*** The student's problem solution applies relevant concepts, skills, strategies, and technology with extreme efficiency and accuracy.
- ***Acceptable:*** The student's problem solution applies relevant concepts, skills, strategies, and technology with moderate efficiency and accuracy.
- ***Inadequate:*** The student's problem solution applies relevant concepts, skills, strategies, and technology with little or no efficiency and accuracy.

I hope you recognize that these quality distinctions—"extreme," "moderate," and "little or no"—are remarkably general and, therefore, almost certain to be interpreted differently by different scorers.

The problem with hypergeneral rubrics is that they sound good, especially on first reading, yet they lack the necessary specificity to let teachers or students know for sure what's going on. A rubric-user, even a diligent one, reads a hypergeneral rubric and discovers that excellent is better than good, good is better than fair, and fair is better than wretched. The excessively general distinctions in a hypergeneral rubric promise more than such a rubric can possibly deliver.

One reason that educators sometimes succumb to the seductive allure of hypergeneral rubrics is that the amalgam of subskills they're trying to assess via a student's response to a single task turns out to be so divergent (in other words, so heterogeneous), that the rubric-developer simply can't come up with meaningful one-size-fits-all evaluative criteria *or* qualitative distinctions. Actually, if that's the case, then the "grain size" of the cognitive skill being taught and tested (that is, the breadth of the skill) is probably too large. If a rubric creator can't contrive anything more useful than a hypergeneral rubric, you can be sure the rubric won't be much good for systematic scoring—because of its considerable ambiguity—nor much good for guiding instruction—because of its considerable ambiguity. Hypergeneral rubrics, if you encounter any, should be recognized for what they are: impostors in an educator's quest for curricular clarity.

Task-Specific Rubrics

A second kind of rotten rubric is one that can only be used when judging a student's response to a particular task. It is called a *task-specific rubric*, and its name gives away its great strength—from a scoring perspective—and its great weakness—from an instructional perspective.

Remember, rubrics are most sensibly used when educators are trying to get a fix on the degree to which a student has mastered a significant skill—usually a cognitive skill. The typical way to get students to demonstrate their level of skill-mastery is to present them with one or more tasks that require them to employ the skill in which you're interested. For example, if you wish to see if a group of science students can employ the "inquiry method" when conducting experimental investigations, you might present a task to students asking them to design, carry out, and report an experiment-based investigation involving, say, earthworms. If students carry out the earthworm experiment successfully, this provides evidence that they seem to have a good grasp of the "inquiry method."

Well, it is possible to develop a rubric so that it can be used *exclusively* to evaluate students' responses to this particular earthworm-task. It is, in every sense, a *task-specific rubric*. Task-specific rubrics, clearly, are remarkably useful when scoring students' responses to a particular task.

These rubrics are essentially worthless, however, when considered instructionally because they deal with a student's response to a particular task—a task that, almost certainly, the student will never again encounter. What teachers really want to do is promote students' mastery of a skill that can be successfully applied to an unlimited number of tasks—not just

one. And a rubric that's riveted on only one limited task will be of no help—to students or teachers—in clarifying the important elements of *the skill itself*. After all, how many earthworm experiments are really needed in life?

Task-Specific Rubrics: Scoring guides suitable for judging responses to only a particular task

Let me try to make this point by showing you, in Figure 4, a pair of excerpts from a rubric that's intended to help evaluate fifth-graders' ability to write narrative essays. Let's say that the teacher has worked up a rubric to appraise students' narrative essays, and one of the rubric's evaluative criteria is *organization*, that is, the way such an essay is put together structurally.

Okay, I want you to imagine that Mrs. Hill, who teaches fifth-graders at Shelby Elementary School, arranged to have a nurse from the nearby Red Cross blood bank visit her class and describe what goes on when citizens donate blood. Just yesterday, the nurse spent an hour with Mrs. Hill's students during which he, in order, (1) described the importance of people's giving blood, (2) set forth the step-by-step sequence of how blood is donated, (3) told students how the recent concerns about "mad cow" disease have changed the way potential blood donors are screened, and (4) recounted how one donor had forgotten an earlier donation and thus tried to give blood only one week after the person's original donation. He then answered students' questions.

The next day, Mrs. Hill asks her students to compose a narrative essay telling the story of the nurse's visit to class. Now, please take a look at how Mrs. Hill, if she were devising a task-specific rubric versus a more generally applicable rubric, might have described the quality distinctions for her rubric's organization criterion. You'll see those two examples in Figure 4.

Figure 4. Two Illustrative Quality Distinctions for an Evaluative Criterion in a Rubric Intended to Evaluate Students' Narrative Writing Skill

A Task Specific Distinction

Evaluative Criterion:

Organization. The best possible demonstration of a student's organizational ability will be seen in essays which, in the order that the following events occurred in class, describe (1) the importance of blood-giving, (2) the steps in a donor's giving blood, (3) the impact of "mad cow" disease, and (4) the incident regarding a forgetful blood

donor. To the extent that an essay omits any of these four topics or distorts the order in which they were presented, lower evaluations of the essay's organization should be given.

A More Widely Applicable Distinction

Evaluative Criterion:

Organization. To receive complete credit on this criterion, an essay must contain a three-component framework, that is, an introduction, a body, and a conclusion. The essay itself must employ a defensible structure in a manner most appropriate for the task itself, for instance, by using an order-of-importance, logical, or chronological structure. Typically, but not always, well organized narrative essays will be based on a chronological sequence. Essays will be evaluated adversely if they do not incorporate a clear introduction, body, and conclusion, or if they employ an indefensible structure.

Notice that the first description in Figure 4, that is, the task-specific one, offers students and teachers no clues about how to cope with the *next* narrative essay-writing task. The rubric is locked on only the one task about the recent visit of the nurse from the blood bank. The bottom description, the more widely applicable one, presents an evaluative structure that can be used when creating narrative essays for any task. The students can learn that such an essay must contain a *three-component framework*, namely, an introduction, a body, and a conclusion. Moreover, a *defensible structure* such as a logical or order-of-importance arrangement must be employed.

I hope you can see that the top excerpt applies solely to a student's *task*, while the bottom excerpt focuses on a student's mastery of a *skill*. And that's the sort of rubric I want to deal with now, the kind of rubric that will be more helpful, on instructional grounds, to both teachers and students.

Skill-Focused Rubrics

Whereas hypergeneral and task-specific rubrics have enormous shortcomings, a *skill-focused rubric*, if sensibly constructed, is ideal for clarifying curricular intentions. A skill-focused rubric, as its name implies, is directed toward the qualities deemed salient when students apply the skill being taught (and tested) to any task for which that skill is applicable.

Skill-Focused Rubrics: Scoring guides for judging students' mastery of the skill being assessed

How to build a skill-focused rubric. I know that this *MA* booklet is only supposed to promote a reader's *understanding* of several rubric-relevant topics and is not supposed to transform readers into roaring rubric-builders. Even so, I think that if I briefly sketch the steps with which you might build a skill-focused rubric, you'll then better understand what such a rubric is and what it is supposed to accomplish. Here, then, are five steps in creating a skill-focused rubric:

- **Step 1.** ***Select a genuinely significant skill to be taught/tested.*** Because of the time it takes to build and use rubrics, be sure to choose a cognitive skill to be taught—and later assessed—that's really a challenging, knock-your-socks-off important skill.

- **Step 2.** ***Choose a modest number of evaluative criteria, all of which students can be taught to master.*** Identify the most meaningful factors by which to judge the quality of a student's performance. Ideally, choose only a small number, and be sure that each of those evaluative criteria can be successfully taught to students.

- **Step 3.** ***Concisely label each evaluative criterion.*** Because the evaluative criteria constitute the key building blocks for students' skill-acquisition, supply succinct explanatory names for each criterion which, with sufficient usage, can promote students' familiarity with the rubric's evaluative structure.

- **Step 4.** ***Spell out the quality distinctions for each evaluative criterion in a manner readily comprehensible to both teachers and students.*** Choose a way to define strong and weak student performances with respect to each evaluative criterion you have incorporated. These distinctions should evoke essentially identical interpretations from all rubric-users.

- **Step 5.** ***Make the rubric as brief as possible without distorting the essence of its evaluative function.*** Busy teachers rarely have time to read lengthy treatises, so the more terse a rubric is, the more likely it will be used by teachers and students. What you're looking for here is sensible succinctness.

One of the best ways to get instructional mileage out of a well-formed rubric is to make sure that students truly understand the rubric's innards. Some teachers report that they have had great success when they actively involve their students in the creation of a rubric. That

> ***involvement, of course, can lead to increased ownership of the resultant rubric. I've also talked to a number of other teachers who indicate that when they have embarked on the generation of student-built rubrics, the end product is far from glorious. I suppose the caliber of any student-created rubrics, unsurprisingly, depends on the particular students' abilities and the teacher's guidance capabilities.***

These, then, are five steps involved in creating skill-focused rubrics. I hope you can see that the mission of such rubrics is to permit accurate scoring of students' responses to a skill-requiring task—using evaluative criteria that are applicable to judging students' responses to any such skill-determining task. Because this sort of rubric tries to do a really good job in explicating the key innards of a significant cognitive skill, its contribution to curricular clarity will be considerable.

In Figure 5 you'll see another illustrative rubric that upper-elementary teachers might use when evaluating their student's oral-reporting skills. As usual, the more conversant students become with the rubric's evaluative criteria, the more apt those students are to employ those factors when judging the caliber of not only their own oral reports but also the oral reports of their classmates.

Figure 5. An Illustrative Rubric for Evaluating Upper-Elementary Students' Oral-Reporting Skill

Skill: Students will be able to make effective 5–10 minute oral reports to classmates of individually conducted research projects.

Evaluative Criteria:

1. *Organization:* *Is the oral report effectively structured?*

High quality reports will contain a clear introduction, body, and conclusion. Low quality reports will lack one or more of these components.

2. *Presentation Style:* *Is the oral report presented in a conversational, relaxed manner?*

High quality reports will see students speaking easily, maintaining effective eye-contact with classmates, and relying on few filler words and sounds (such as "uh"). Low quality reports will violate one or more of these three presentation-style considerations.

3. *Content:* *Does the content of the oral report match the student's written report?*

High quality reports will accurately present the most important content of the written report of the student's research investigation. Low quality reports will either fail to include key content in the student's written report and/or inaccurately present such content.

Application Strategy: This rubric can be employed either holistically or analytically. For improvement of young students' oral presentation skills, an analytic application is preferable.

The Figure 5 illustrative rubric, as you can see, relies on only three evaluative criteria. Because students will have been informed that their written reports of a research project must be followed by a 5–10 minute oral report, a skillful teacher will urge students to become familiar with this rubric very early as they carry out their research.

JUDGING RUBRICS

Varied Formats, Qualitative Differences

By now I hope you recognize that there is no single format that must mandatorily be followed when rubrics are created. The rubric-components I've been describing can be presented with dramatically different labels and, of course, arranged on pages in all sorts of diverse ways.

However, what I want you to be on the watch for, either in the rubrics you develop yourself or in rubrics developed by others, is the essential nature of those rubrics. Are they, at bottom, too general and vague to really do you any good? Or are they so directly based on a particular task that they have no widespread applicability? Or, hopefully, are they focused directly on the key features of the cognitive skill that students are supposed to be mastering? Remember my earlier admonition that all rubrics should not be considered worthwhile. If you find yourself dealing with repugnant rubrics, send them to the nearest recycling receptacle.

A Rubric to Evaluate Rubrics

With all the attention that's been lavished on rubrics in the preceding pages, I would be remiss if I didn't close out this booklet with a rubric you can use to evaluate your own rubrics or the rubrics of others. Accordingly, in Figure 6 I've provided such a rubric. I suppose that, technically, it could be described as a *Rubric Rubric*. But, unless you want to be accused of repeating yourself, I suggest you use another descriptor if you speak of it to your colleagues.

Figure 6. A Rubric for Evaluating Rubrics

When appraising a rubric from an assessment *and* instructional perspective, the evaluative criteria given below should be used. Each criterion is accompanied by a two-directional quality definition. If desired, numerical values can be assigned when applying the criteria. The rubric can be used analytically or holistically.

Evaluative Criteria:

1. *Significance.* *Is the skill being assessed a genuinely worthwhile one?*

A strong rubric will be focused on students' attainment of a high-level cognitive skill that requires meaningfully lengthy instruction to promote. A weak rubric will focus on students' acquisition of knowledge or a quickly taught, low-level cognitive skill.

2. *Evaluative Criteria.* *Have the rubric's scoring criteria been selected so they are few in number, succinctly labeled, and instructionally addressable?*

A strong rubric will contain only a modest number of concisely labeled evaluative criteria, each of which students can be taught to employ when appraising their own mastery of the skill being assessed. A weak rubric will contain too many poorly labeled (or unlabeled) evaluative criteria, some of which students cannot be directly taught to employ.

3. *Quality Distinctions.* *Are degrees of excellence satisfactorily described for each of the rubric's evaluative criteria?*

A strong rubric will provide sufficiently clear descriptions of qualitative differences in how each evaluative criterion is applied so that, with reasonable training, different

rubric-users would be able to apply those criteria in essentially the same way. A weak rubric's evaluative criteria will be accompanied by qualitative differentiations that lead to diverse interpretations.

4. *Concise Clarity.* *Is the rubric presented in a sufficiently succinct and lucid manner so that it is likely to be used?*

A strong rubric will, given its important function in delineating an evaluative process, be presented briefly enough and clearly enough so that busy teachers are apt to use it. A weak rubric will be too lengthy or too technical for its intended users.

You'll note that I've set forth four evaluative criteria in Figure 6 to use when appraising a rubric. I'm sure that there would be differences if other individuals came up with their own rubric to evaluate rubrics. For me, though, I usually judge a rubric on the basis of (1) the significance of the cognitive skill that the rubric is being used to assess, (2) the caliber of the rubric's evaluative criteria, (3) the way that the rubric lays out the qualitative differences for the application of its evaluative criteria, and (4) the brevity and clarity of the rubric.

I'm sure you recognize that there's a considerable degree of artistry required to generate a truly excellent rubric. For instance, the fourth evaluative criterion in Figure 6 is *concise* clarity. If a rubric-designer had no limits on the amount of verbiage that could be shoveled into a rubric, then clarity could surely be attained simply by ladling out wads of words at the rubric-user. But unused rubrics don't help teachers and, therefore, don't help students.

Remember, the overriding rule of a rubric is to promote greater clarity regarding how a teacher is going to judge students' skill-acquisition. And that clarity, if relied on by teachers when designing and delivering instruction, will benefit both the testing *and* the teaching of students.

RECAP AND WRAP-UP

Back at the beginning of this booklet, I confessed in advance that I wanted you to understand:

- ***What a rubric is and why it can be helpful in both testing and teaching.***
- ***What the three essential components of a rubric are.***
- ***How to distinguish between instructionally charming and instructionally churlish rubrics.***

I hope that, for the most part, you've now picked up those three yearned-for outcomes. I realize all too well that your attainment of such understandings will not automatically transform you into a remarkably good rubric-writer. After all, that's a skill, and, as you surely know, the acquisition of a skill usually takes plenty of practice and—more frequently than we'd prefer—some serious stumbling. Candidly, I think that if I put my mind to it these days, I can usually crank out a pretty fair rubric. But I hesitate to tell you how many flawed rubrics I've written. There have been many, many, many of those.

But I'm convinced that the sets of understandings you hopefully have picked up as you roamed through this booklet will put you into a position—if you wish—to begin developing your own rubrics. And if you do construct a rubric, try to improve it by getting a colleague to react to it—but, of course, only a colleague who's already completed this booklet!

At the very least, you should apply your new understandings to appraising the caliber of any rubric that you find yourself being asked to use with your students. One of this booklet's key points is that some rubrics help and some rubrics don't. You should now be in a position to applaud skill-focused rubrics while sneering malevolently at task-specific or hypergeneral rubrics.

Rubrics, as I've tried to contend more than once in these pages, can be a marvelous ally of teachers and students. But to make the kind of contribution that rubrics really are able to, a rubric must be first-rate. I hope that by now you can scrutinize a rubric well enough to tell whether that rubric should be shipped off to a school or sent to a waste-disposal plant.

GLOSSARY TERMS

Analytic Scoring: When a rubric's evaluative criteria are applied separately in judging students' work

Application Strategy: Use of a rubric's evaluative criteria holistically (all together) or analytically (one-by-one)

Cognitive Skill: A specific intellectual competency that individuals can apply in appropriate situations

Constructed-Response Items: Test items requiring students to generate their answers

Evaluative Criteria: Factors considered when using a rubric to judge a student's response

Holistic Scoring: When a rubric's evaluative criteria contribute to a single, overall judgment of quality

Hypergeneral Rubrics: Excessively inexact, often vague, scoring guides

Knowledge: Memorized information, for example, recollectable facts, dates, principles, and terminology

Performance Test: An assessment task typically requiring students to create a somewhat elaborately constructed response

Quality Distinctions: Descriptions of different qualitative levels for each of a rubric's evaluation criteria

Rubric: A scoring guide used to judge students' work

Selected-Response Items: Test items requiring students to choose from two or more options

Skill-Focused Rubrics: Scoring guides for judging students' mastery of the skill being assessed

Student Behavior: A student's performance intended to demonstrate what the student has learned

Student Product: Something material a student creates to demonstrate the student's learning

Task-Specific Rubrics: Scoring guides suitable for judging responses to only a particular task

REFERENCES

Arter, Judy and Jan Chappuis. *Developing and Recognizing Quality Rubrics.*
Portland, OR: Assessment Training Institute, 2006.

Bloom, Benjamin S., et al. *Taxonomy of Educational Objectives: Handbook 1: Cognitive Domain.*
New York: David McKay, 1956.

Stevens, Dannelle D. and Antonia J. Levi. *Introduction to Rubrics: An Assessment Tool to Save Grading Time, Convey Effective Feedback and Promote Student Learning.*
Herndon, VA: Stylus Publishing, 2004.

Booklet design: Scott Santoro, Worksight

Published in 2006 by
Routledge
Taylor & Francis Group
270 Madison Avenue
New York, NY 10016

Published in Great Britain by
Routledge
Taylor & Francis Group
2 Park Square
Milton Park, Abingdon
Oxon OX14 4RN

Routledge is an imprint of Taylor & Francis Group

Printed in the United States of America on acid-free paper
10 9 8 7 6 5 4 3 2 1

International Standard Book Number-10: 0-415-95240-9 (Softcover)
International Standard Book Number-13: 978-0-415-95240-8 (Softcover)

Library of Congress Cataloging-in-Publication Data

Catalog record is available from the Library of Congress

informa
Taylor & Francis Group
is the Academic Division of Informa plc.

Visit the Taylor & Francis Web site at
http://www.taylorandfrancis.com

and the Routledge Web site at
http://www.routledge-ny.com