

W. James Popham

What's Wrong—and What's Right—with Rubrics

Rubrics have the potential to make enormous contributions to instructional quality—but first we have to correct the flaws that make many rubrics almost worthless.

Rubrics are all the rage these days. It's difficult to attend an educational conference without running into relentless support for the educational payoffs of rubrics. Indeed, the term itself seems to evoke all sorts of positive images. Rubrics, if we believe their backers, are incontestably good things.

But for many educators, rubrics inspire a series of questions. What are rubrics, and where did they come from? What is an educationally appropriate role for rubrics? Why do so many current rubrics fail to live up to their promise as guides for both teachers and students? What should we do to make rubrics better?

The Rudiments of Rubrics

As used today, the term *rubric* refers to a scoring guide used to evaluate the quality of students' constructed responses—for example, their written compositions, oral presentations, or science projects.

A rubric has three essential features: evaluative criteria, quality definitions, and a scoring strategy.

Evaluative criteria are used to distinguish acceptable responses from unacceptable responses. The criteria will obviously vary from rubric to rubric, depending on the skill involved. For instance, when evaluating written compositions, teachers often use such evaluative criteria as organization, mechanics, word choice, and supporting details. Evaluative criteria can either be given equal weight or be weighted differently.

Quality definitions describe the way that qualitative differences in students' responses are to be judged. For instance, if *mechanics* is an evaluative criterion, the rubric may indicate that to earn the maximum number of points for mechanics, a student's composition should contain no mechanical errors. The rubric must provide

a separate description for each qualitative level. This means that if four different levels of quality are assigned to a written composition's organization, the rubric provides descriptions for each of those levels.

A *scoring strategy* may be either holistic or analytic. Using a holistic strategy, the scorer takes all of the evaluative criteria into consideration but aggregates them to make a single, overall quality judgment. An analytic strategy requires the scorer to render criterion-by-criterion scores that may or may not ultimately be aggregated into an overall score.

The Roots of Rubrics

The original meaning of *rubric* had little to do with the scoring of students' work. The *Oxford English Dictionary* tells us that in the mid-15th century, *rubric* referred to headings of different sections of a book. This stemmed from the work of Christian monks who painstakingly reproduced sacred literature, invariably initiating each major section of a copied book with a

large red letter. Because the Latin word for red is *ruber*, *rubric* came to signify the headings for major divisions of a book.

A couple of decades ago, *rubric* began to take on a new meaning among educators. Measurement specialists who scored students' written composi-

tions began to use the term to describe the rules that guided their scoring. They could have easily employed a more readily comprehensible descriptor, such as *scoring guide*, but *scoring guide* lacked adequate opacity. *Rubric* was a decisively more opaque, hence technically attractive, descriptor.

A Rubric's Role

Typically, people don't use rubrics unless the constructed response being judged is fairly significant. Thus, teachers rarely use rubrics to judge students' responses on short-answer tests; and, of course, rubrics are unnecessary for scoring tests like multiple-choice exams. With a few exceptions, teachers use rubrics to judge the adequacy of students' responses to performance tests.

The vast majority of rubrics are masquerading as contributors to instruction when, in reality, they have no educational impact at all.

Many rubrics now being billed as instructionally useful
provide teachers and students with absolutely
no cues about what is genuinely significant in
a student's response.

A performance test presents a demanding task to a student, then asks the student to respond to the task in writing, orally, or by constructing some type of product—for example, composing a persuasive essay on a given topic. Educators ordinarily use performance tests when they want to determine a student's status with respect to a significant skill. Based on the student's level of achievement on a performance test, educators make an inference about the degree to which the student has mastered the skill the test represents. Excellent results on the performance test imply that the student has mastered the skill; poor results suggest the opposite.

Because performance tests typically call for students to display fairly high-level skills and because the tasks involved are often *authentic* (that is, they resemble real-world challenges), performance tests have received substantial support from educators and noneducators alike. The subsequent increased use of performance tests has made rubrics popular—students' responses have to be scored. Consequently, most commercial textbook publishers are creating rubrics for their end-of-chapter tests, and the testing firms that distribute and score standardized achievement tests are introducing rubrics into their scoring operations.

Performance tests are intended to measure students' mastery of important skills—those that educators regard as worth promoting instructionally. Why, indeed, should anyone go to the trouble of building a performance test to measure students' mastery of a trivial skill or an innate attribute that's impervious to instruction? Instructors seek to enhance students' skill mastery. If performance tests are truly worth the effort that goes into creating and using

them, we should evaluate them chiefly according to the contributions they make to students' skill mastery.

What's Wrong with Rubrics?

Although rubrics are receiving near-universal applause from educators, the vast majority of rubrics are instructionally fraudulent. They are masquerading as contributors to instruction when, in reality, they have no educational impact at all. Here are four flagrant flaws that are all too common in teacher-made and commercially published rubrics.

Flaw 1: Task-specific evaluative criteria. A rubric's most important component is the set of evaluative criteria to be used when judging students' performances. The criteria should be the most instructionally relevant component of the rubric. They should guide the teacher in designing lessons because it is students' mastery of the evaluative criteria that ultimately will lead to skill mastery. Moreover, teachers should make the criteria available to students to help them appraise their own efforts.

But what if the evaluative criteria in a rubric are linked only to the specific elements in a particular performance test? Unfortunately, I've run into a flock of such task-specific rubrics these days, especially in the most recent crop of nationally standardized tests that call for constructed responses from students.

Consider, for example, a task that presents a cross-section picture of a vacuum bottle, then calls on students to identify the materials that had to be invented before vacuum bottles could be widely used. Such tasks are interesting, often inventive, and may even be fun for students to do. But the accompanying rubric has evaluative criteria that are totally task-specific. Each criterion is linked to the students' proper interpre-

tation of the features of the picture that accompanies the test item. Each is exclusively based on a specific task in a single performance test.

How can such task-specific criteria help guide a teachers' instructional planning? How can they help students evaluate their own efforts? Perhaps the commercial test publishers are eager to install task-specific evaluative criteria because such criteria permit more rapid scoring with a much greater likelihood of between-scorer agreement. But such criteria, from an instructional perspective, are essentially worthless. Teachers need evaluative criteria that capture the essential ingredients of the skill being measured, not the particular display of that skill applied to a specific task.

Flaw 2: Excessively general evaluative criteria. Just as task-specific evaluative criteria render a rubric instructionally useless, so too do excessively general evaluative criteria. Numerous rubrics have criteria so amorphous they are almost laughable.

Many commercially published rubrics provide several qualitative levels so that teachers can ostensibly distinguish among students' performances. The highest level of student performance is labeled "advanced"—or some suitable synonym, then described as "a superior response to the task presented in the performance test—a response attentive not only to the task's chief components, but also its nuances." A second, lower level of response is described in slightly less positive terms, and so on. In essence, these overly general criteria allow both teachers and students to conclude that really good student responses to the task are, well, really good. And, of course, really bad student responses are—you guessed it—really bad.

I'm exaggerating a bit—but not much. Many rubrics now being billed as

instructionally useful provide teachers and students with absolutely no cues about what is genuinely significant in a student's response, and they offer teachers no guidance on the key features of the tested skill.

Flaw 3: Dysfunctional detail.

Another shortcoming in many rubrics is excessive length: Busy teachers won't have anything to do with them. If we want rubrics to make a difference in classroom instruction, we need to create rubrics that teachers will use. Lengthy, overly detailed rubrics are apt to be used only by inordinately compulsive teachers.

Many of the rubrics being circulated these days are lengthy and laden with details. After all, most of the earliest rubrics were created for use in large-scale, high-stakes assessments. If a state's high school diploma were to be based on how well a student functioned on an important statewide performance test—a writing sample, for instance—the architects of the accompanying rubric understandably might have leaned toward detailed scoring rules. In general, the more detailed and constraining a rubric's scoring rules, the greater the likelihood of between-rater agreement. For high-stakes tests, detailed rubrics were common.

When educators and textbook publishers introduced rubrics for classroom use, many models came from these earlier large-scale assessments. But such lengthy, excessively detailed rubrics almost invariably turn teachers off—an unfortunate effect, because a properly fashioned rubric can really improve the caliber of instructional activities.

In contrast to a brief rubric, detailed rubrics will, of course, spell out more precisely how to ascertain the quality of a student's response. A one- or two-page

Rubric developers should remember that their efforts should guide teachers, not overwhelm them.

rubric will be subject to wider interpretation than will a six-page, "lay out all the scoring rules" rubric. But the practical choice comes down to this: (1) Should we build short rubrics that offer less than stringent scoring guidance but will be used by teachers? or (2) Should we build lengthier rubrics that provide stringent scoring guidance but won't be used?

Happily, in almost all instances, lengthy rubrics probably can be

reduced to succinct but far more useful versions for classroom instruction. Such abbreviated rubrics can still capture the key evaluative criteria needed to judge students' responses. Lengthy rubrics, in contrast, will gather dust.

Flaw 4: Equating the test of the skill with the skill itself. This problem stems less from rubrics themselves than from an error made by rubric users. A particularly prevalent misunderstanding occurs when rubric users become so caught up with the particulars of a given performance test that they begin thinking of the test as the skill itself. For example, if the performance test calls for a student to display mathematical problem-solving skills by carrying out a specific multi-step solution, far too many teachers become fixated on the student's

F I G U R E 1

A Rubric That Improves Instruction

The following mathematical task includes three subtasks, for which rubrics would be an appropriate aid to instruction and assessment:

Task: Present students with reality-based raw data, then ask them to (1) compute several averages, (2) present those averages in a prescribed graphic form, and (3) draw a defensible conclusion from the graphed averages.

For the graphing subtask, we have identified one or more evaluative criteria and have used three levels of quality in scoring students' responses, with three points being the highest score. To illustrate the levels of quality, teachers can use examples of student work from previous years.

Analytic Scoring Rubric for Subtask 2: Graphing (Evaluative criteria: accuracy, quality of title, and quality of axis and interval labels.)

Highly Proficient (3 points): Student has constructed a completely *accurate* task-prescribed graph (for example, bar, pie, or line graph), and the *title*, *axis*, and *interval labels* are all appropriate.

Proficient (2 points): Student has constructed an almost completely *accurate* task-prescribed graph (for example, bar, pie, or line graph), and the *title*, *axis*, and *interval labels* are almost all appropriate.

Not Yet Proficient (1 point): Student has *not accurately* constructed a task-prescribed graph (for example, bar, pie, or line graph), and fewer than half the *title*, *axis*, and *interval labels* are appropriate.

Source: Based on a rubric developed by Jeanne Miyasaka, WestED, 2221 E. Turquoise, Phoenix, AZ 85028.

mastery of that particular multistep solution as the aim of instructional efforts. These teachers strive for test mastery rather than skill mastery.

Realistically, any really worthwhile skill can probably be measured by an array of tasks that could be embodied in different performance tests. For example, to determine a student's ability to give an extemporaneous speech, a teacher might allow a student to choose from many topics for the "speech" performance test. As a practical matter, of course, teachers don't have time for students to display a single skill through a dozen different performance tests. Although the more performance tests a student completes, the more accurate will be the inferences about skill mastery, teachers usually rely on a single performance test.

Nevertheless, teachers must instruct toward the *skill* represented by the performance test, not toward the *test*. Test-focused instruction, especially if it mimics the test in every detail, will often stifle the student's general mastery of the skill. Students may, indeed, learn how to do well on a given performance test, but if asked to tackle a different performance test—a test derived from the same skill—they may stumble. Teachers must keep in mind that performance tests represent skills. The tests are not the skills themselves.

Getting Rubrics Right

Having maligned many of today's rubrics, it's time to get constructive. What would a rubric look like that not only helped teachers judge the quality of students' responses to a performance test but also assisted those teachers in helping students acquire the skill represented by that test?

For openers, such a rubric would contain three to five evaluative criteria.

Rubrics are not only scoring tools but also, more important, instructional illuminators.

It is tempting to lay out all of the possible criteria that could be used to judge students' responses; but rubric developers should remember that their efforts should guide teachers, not overwhelm them. In rubrics, less is more.

Second, each evaluative criterion must represent a key attribute of the skill being assessed. Each criterion must be teachable in the sense that teachers can help students increase their ability to use the criterion when tackling tasks that require that skill. For example, many teachers are quite competent in helping students learn how to compose essays that embody skillful organization, effective word choice, appropriate mechanics, and suitable supporting detail. Each of these criteria is eminently teachable. Effective teachers of composition will share these criteria with students to help them master essential writing skills.

Figure 1 presents a brief, instructionally oriented rubric for a mathematics skill requiring students to complete three subtasks: averaging, graphing, and concluding. As the "graphing" rubric shows, each subtask has teachable evaluative criteria, and those criteria are applicable across a wide range of similar subtasks. This rubric does not delineate the nuances of each evaluative criterion so that different people using the rubric would invariably score students' responses in an identical manner. But if you are writing a rubric and are faced with a choice between interscorer

agreement and instructional impact, opt for the latter.

I do not want to suggest that the isolation of teachable evaluative criteria for rubrics is fools' play. It isn't. But by now I have seen enough rubrics containing teachable evaluative criteria that I am confident such rubrics can be created.

The more quickly we abandon both task-specific and excessively general rubrics, the more likely we will come up with rubrics that actually enhance instruction. In addition, for routine use, relatively short rubrics must be the rule. If we want teachers to focus their instructional attention on the evaluative criteria embedded in rubrics, rarely should a rubric exceed one or two pages. With any rubric intended for classroom use, a sheaf of papers held by a staple should be regarded as an enemy.

Rubrics represent not only scoring tools but also, more important, instructional illuminators. Appropriately designed rubrics can make an enormous contribution to instructional quality. Unfortunately, many rubrics now available to educators are not instructionally beneficial. If these flawed rubrics are not rapidly replaced with instructionally helpful ones, then the educational promise of rubrics will surely not be realized. ■

Author's note: Write to the author for copies of other rubrics regarded as exemplary from an instructional perspective. See also the December 1996–January 1997 issue of *Educational Leadership*.

W. James Popham is Emeritus Professor at the University of California-Los Angeles and Director of IOX Assessment Associates, 5301 Beethoven St., Suite 208, Los Angeles, CA 90066-7061 (e-mail: wpopham@ucla.edu).

