Rubric for Rubrics

Overview

The *Rubric for Rubrics* is a rubric for evaluating the quality of performance assessment rating scales, which we call *rubrics* or *scoring guides*. It describes the features of a rubric that make it useful for assessment and learning in the classroom. It is intended to be used with general, not task-specific rubrics.

We developed the *Rubric for Rubrics* to evaluate rubrics for use in the classroom, not for use with large-scale assessments such as state or provincial assessments. Although many features of quality would be the same for both uses, large-scale rubrics often end up with features that would be counterproductive in a rubric intended for classroom use. For example, developers of rubrics for large-scale uses frequently emphasize a quick, overall picture of student performance—no detail. Rubrics used in the classroom, on the other hand, often need to provide detailed diagnostic information to inform day-to-day instructional decisions.

Definitions:

- A *criterion* is a key dimension of quality useful to consider separately. The *Rubric for Rubrics* has two criteria: *Coverage/Organization* and *Clarity*.

- Subheads under each criterion are *indicators*, and the following numbered items are *descriptors*. An indicator, for example is, "Covers the right content." A descriptor for that indicator is, "The content of the rubric represents the best thinking in the field . . ." There are three indicators for the first criterion and two for the second criterion.

- *Levels* are points on a rating scale defining degrees of quality. There are typically three to six levels of performance on rubrics. The *Rubric for Rubrics* has five levels, ranging from *Strong* to *Weak.*

Content of the *Rubric for Rubrics*

Criterion 1: Coverage/Organization

The *content* of a classroom rubric defines what to look for in a student's product or performance to determine its quality; what will "count." Teachers and students use this content to determine what they must do in order to succeed. What students see is what you'll get. If the rubric has problems on this criterion, there is no need to continue to the *Clarity* criterion. There are three indicators for the first criterion.

Indicator 1A: Covers the Right Content

A classroom rubric should (1) bear a direct relationship to the content standards and learning targets it is intended to measure, (2) cover all essential features that create quality in a product or performance, (3) leave out all trivial or unrelated features, and (4) support and extend your understanding about what you actually *do* look for when evaluating student work.

Indicator 1B: Criteria Are Well Organized

The list of features that describe quality should be as concise as possible and organized into a usable form. This often involves identifying and grouping similar features into criteria and being sure that the relative importance given to each criterion represents its relative contribution to the quality of the product or performance as a whole.

Indicator 1C: Number of Levels Fits Targets and Uses

The number of levels needs to be appropriate for the intended learning target and your use of the rubric. There needs to be enough levels to track student progress without so many that users can't distinguish among them.

Criterion 2: Clarity

A classroom rubric is *clear* to the extent that teachers, students, and others are likely to interpret the statements and terms in the rubric the same way. A rubric can be strong on the criterion of *Coverage/Organization* but weak on the criterion of *Clarity*—the rubric seems to cover the important dimensions of performance, but doesn't describe them very well. Likewise, a rubric can be strong on the criterion of *Clarity,* but weak on the criterion of *Coverage/Organization*—it's very clear what the rubric means, but it is not focused on the right criteria. There are two indicators for the criterion of *Clarity*.

Indicator 2A: Levels Defined Well

The key with *Clarity* is to define levels so transparently that students (and teachers) can see precisely what features of work cause people to agree that work is Strong, Medium, or Weak. The instructional usefulness of any rubric depends on the clarity of level descriptions.

Indicator 2B: Levels Parallel

Rubrics should include a parallel feature of work on each level. For example, if you find that a rubric for playing the violin contains "lackadaisical bowing" as one descriptor of a middle-level performance, then a statement about the quality of the bowing must be included at the Strong and Weak levels as well. If this descriptor is not referred to at other levels, the levels are not parallel.

How to Use the Rubric

The descriptors under each indicator are not meant to function as a checklist. Rather, they are meant to help users determine the level of quality of the classroom rubric under consideration. Not everything has to be present (or missing) for the classroom rubric to be judged to be at a particular level of quality. Ask yourself, "Which level of descriptors best describes the classroom rubric I'm considering?"

An odd number of levels is used because the middle level represents a balance of strengths and weaknesses. It would take some work to make it usable, but it probably is worth the effort. A Strong score doesn't necessarily mean that the classroom rubric under consideration is perfect; rather, it means that it would require very little work to get it ready for use. A Weak score means that the classroom rubric needs so much work that it probably isn't worth the effort—it's time to find another one. It might even be easier to begin from scratch.

Additionally, a Medium score does not mean *average.* This is a criterion-referenced scale, not a norm-referenced one. It is meant to describe levels of quality in a classroom rubric, not to compare those currently available. It could be that the typical currently available classroom rubric is closer to Weak than to Medium.

Although three levels are defined, it is in fact a *five-level* scale. Think of level 4 as a combination of characteristics from levels 5 and 3. Likewise, level 2 combines characteristics from levels 3 and 1.

Criterion 1: COVERAGE/ORGANIZATION		
A. Covers the Right Content		
5—Strong	*3—Medium*	*1—Weak*
1. The content of the rubric represents the best thinking in the field about what it means to perform well on the skill or product under consideration.	1. Much of the content represents the best thinking in the field, but there are a few places that are questionable.	1. You can't tell what learning target(s) the rubric is intended to assess, or you can guess at the learning targets, but they don't seem important, or content is far removed from current best thinking in the field about what it means to perform well on the skill or product under consideration.
2. The content of the rubric aligns directly with the content standards/ learning targets it is intended to assess.	2. Some features don't align well with the content standards/learning targets it is intended to assess.	2. The rubric doesn't seem to align with the content standards/learning targets it is intended to assess.
3. The content has the "ring of truth"—your experience as a teacher confirms that the content is truly what you do look for when you evaluate the quality of a student performance or product. In fact, the rubric is insightful; it helps you organize your own thinking about what it means to perform well.	3. Much of the content is relevant, but you can easily think of some important things that have been left out or that have been given short shrift, or it contains an irrelevant criterion or descriptor that might lead to an incorrect conclusion about the quality of student performance.	3. You can think of many important dimensions of a quality performance or product that are not in the rubric, or content focuses on irrelevant features. You find yourself asking, "Why assess this?" or "Why should this count?" or "Why should students have to do it this way?"

Criterion 1: COVERAGE/ORGANIZATION (Continued)		
B. Criteria are Well Organized		
5—Strong	*3—Medium*	*1—Weak*
1. The rubric is divided into easily understandable criteria as needed. The number of criteria reflects the complexity of the learning target. If a holistic rubric is used, it's because a single criterion adequately describes performance.	1. The number of criteria needs to be adjusted a little: either a single criterion should be made into two criteria, or two criteria should be combined.	1. The rubric is holistic when an analytic one is better suited to the intended use or learning targets to be assessed; or the rubric is an endless list of everything; there is no organization; the rubric looks like a brainstormed list.
2. The details that are used to describe a criterion go together; you can see how they are facets of the same criterion.	2. Some details that are used to describe a criterion are in the wrong criterion, but most are placed correctly.	2. The rubric seems "mixed up"—descriptors that go together don't seem to be placed together. Things that are different are put together.
3. The relative emphasis on various features of performance is right—things that are more important are stressed more; things that are less important are stressed less.	3. The emphasis on some criteria or descriptors is either too small or too great; others are all right.	3. The rubric is out of balance—features of more importance are emphasized the same as features of less importance.
4. The criteria are independent. Each important feature that contributes to quality work appears in only one place in the rubric.	4. Although there are instances when the same feature is included in more than one criterion, the criteria structure holds up pretty well.	4. Descriptors of quality work are represented redundantly in more than one criterion to the extent that the criteria are really not covering different things.
C. Number of Levels Fits Targets and Uses		
5—Strong	*3—Medium*	*1—Weak*
1. The number of levels of quality used in the rating scale makes sense. There are enough levels to be able to show student progress, but not so many levels that it is impossible to distinguish among them.	1. Teachers might find it useful to create more levels to make finer distinctions in student progress, or to merge levels to suit the rubric's intended use. The number of levels could be adjusted easily.	1. The number of levels is not appropriate for the learning target being assessed or intended use. There are so many levels it is impossible to reliably distinguish between them, or too few to make important distinctions. It would take major work to fix the problem.

Criterion 2: CLARITY		
A. Levels Defined Well		
5—Strong	*3—Medium*	*1—Weak*
1. Each score point (level) is defined with indicators and/or descriptors. *A plus:* There are examples of student work that illustrate each level of each trait.	1. Only the top level is defined. The other levels are not defined.	1. No levels are defined; the rubric is little more that a list of categories to rate followed by a rating scale.
2. There is enough descriptive detail in the form of concrete indicators, adjectives, and descriptive phrases that allow you to match a student performance to the "right" score. *A plus:* If students are to use the rubric, there are student-friendly versions, and/or versions in foreign languages for ELL students.	2. There is some attempt to define terms and include descriptors, but some key ideas are fuzzy in meaning.	2. Wording of the levels, if present, is vague or confusing. You find yourself saying such things as, "I'm confused," or "I don't have any idea what this means." Or, the only way to distinguish levels is with words such as *extremely*, *very*, *some*, *little*, and *none*; or *completely*, *substantially*, *fairly well*, *little*, and *not at all*.
3. Two independent users, with training and practice, assign the same rating most of the time. *A plus:* There is information on rater agreement rates that shows that raters can exactly agree on a score 65% of the time, and within one point 98% of the time.	3. You have a question whether independent raters, even with practice, could assign the same rating most of the time.	3. It is unlikely that independent raters could consistently rate work the same, even with practice.
4. If counting the number or frequency of something is included as an indicator, changes in such counts really are indicators of changes in quality.	4. There is some descriptive detail in the form of words, adjectives, and descriptive phrases, but counting the frequency of something or vague quantitative words are also present.	4. Rating is almost totally based on counting the number or frequency of something, even though quality is more important than quantity.
5. Wording is descriptive, not evaluative.	5. Wording is mostly descriptive of the work, but there are a few instances of evaluative labels.	5. Wording tends to be evaluative rather than descriptive of the work; e.g., work is "mediocre," "above average," or "clever."
B. Levels Parallel		
5—Strong	*3—Medium*	*1—Weak*
1. The levels of the rubric are parallel in content—if an indicator of quality is discussed in one level, it is discussed in all levels. If the levels are not parallel, there is a good explanation why.	1. The levels are mostly parallel in content, but there are some places where there is an indicator at one level that is not present at the other levels.	1. Levels are not parallel in content and there is no explanation of why, or the explanation doesn't make sense.

Classroom Rubric Analysis Form

Rubric: _____

_____

Criterion	Indicator	Rating	Rationale (use words and phrases from the Rubric for Rubrics)
Coverage/ Organization	1A: Covers the Right Content		
	1B: Criteria Are Well Organized		
	1C: Number of Levels Fits Targets and Uses		
Clarity	2A: Levels Defined Well		
	2B: Levels Parallel		