

## HOW TO CREATE A RUBRIC FROM SCRATCH

From Chicago Public Schools 2000: [http://www.intranet.cps.k12.il.us/assessments/Ideas\\_and\\_Rubrics](http://www.intranet.cps.k12.il.us/assessments/Ideas_and_Rubrics)

Sometimes it is not possible to find an appropriate existing rubric to use or modify. The purpose of this section is to explain how to develop a scoring rubric "from scratch." The steps you will need to follow are shown in the box below.

### OVERVIEW OF STEPS IN DEVELOPING A SCORING RUBRIC

- Step 1. Brainstorm a list of dimensions or key attributes of the product or performance or product to be assessed.
- Step 2. Look at some actual examples of student work to see if you have omitted any important dimensions.
- Step 3. Refine and consolidate your list of dimensions as needed.
- Step 4. Write a definition of each of the dimensions.
- Step 5. Develop a continuum (scale) for describing the range of products/performances on each of the dimensions.
- Step 6. Alternatively, instead of a set of rating scales, you may choose to develop a holistic scale or a checklist on which you will record the presence or absence of the attributes of a quality product/performance.
- Step 7. Evaluate your rubric using the criteria discussed in Part 1.
- Step 8. Pilot test your rubric or checklist on actual samples of student work.
- Step 9. Revise the rubric and try it out again.
- Step 10. Share the rubric with your students and their parents.

\*\*\*\*\*

### STEP 1: BRAINSTORM A LIST OF DIMENSIONS OR KEY ATTRIBUTES OF THE PRODUCT OR PERFORMANCE OR PRODUCT TO BE ASSESSED.

Good quality performance criteria help teachers answer questions like

"What is expected?"

"What are our standards?"

"What does good performance look like?"

"What do I want to accomplish?" and

"What kind of feedback do I give to improve student work next time?"

-- (Arter, 2000)

Sound criteria "truly reflect the nature of quality in a product or performance. If the criteria leave important things out, include the trivial, or reward performance on features that do not really define quality, then students will learn to the incorrect target."

-- Arter 2000

Brainstorm a list of **dimensions** or **key attributes**. Work with colleagues. Later, work with students. Use guide questions such as:

- What really counts?
- What am I looking for?
- What do I look for when I grade?
- Which elements do I emphasize in teaching?

An example from outside the classroom will illustrate how a scoring rubric is developed. Suppose you were developing a scoring rubric for children's playgrounds. A group brainstorming some possible dimensions for evaluating a playground might come up with a list that looks like this:

- \* whether it's fun
- \* kinds of equipment
- \* safety
- \* places for parents to sit
- \* aesthetically pleasing
- \* cleanliness
- \* well maintained
- \* whether there's graffiti
- \* enough swings
- \* both baby swings and regular swings
- \* variety of equipment
- \* soft surfaces beneath equipment
- \* colorful equipment
- \* sturdy equipment
- \* picnic area
- \* equipment for older and younger children
- \* how attractive
- \* slides
- \* climbing equipment
- \* how the grass looks
- \* how much garbage there is
- \* enough garbage cans
- \* big kids' equipment separate from little kids'
- \* whether equipment looks good, fits the site
- \* equipment allows kids to use imagination

#### OTHER EXAMPLES

- Hawaii students' writing assessments are scored for meaning, voice, clarity, design, and conventions.
- A musical performance might be rated for pitch/harmony, rhythm/tempo, dynamics, tone quality, technique, and style.
- A service learning project can be rated for the academic skills learned, the significance of the service performed for others (did it make a difference?), the initiative and planning shown by students, how well the group worked together, the student's awareness of self as the service provider, and the overall success of the project.

The dimensions you choose may be guided by national curriculum frameworks, publications of professional organizations, sample scoring rubrics (if available), or experts in the subject area in which you are working. The Hawaii Department of Education Standards Toolkits (Curriculum Frameworks, Grade Level Performance Indicator Progression, Scope and Sequence, and Instructional Guides) has some suggestions for dimensions that can be used to assess student work. Alternatively, you and your colleagues may brainstorm a list of as many key attributes of the product/performance to be rated as you can. What do you look for when you grade assignments of this nature? When you teach, which elements of this product/performance do you emphasize?

#### **STEP 2: LOOK AT SOME ACTUAL EXAMPLES OF STUDENT WORK TO SEE IF YOU HAVE OMITTED ANY IMPORTANT DIMENSIONS.**

“Scoring rubrics cannot be developed through professional reflection alone. Rather, that reflection must be coupled with the careful and simultaneous examination of actual samples of relevant student performance. When the two are used in combination, very rich and appropriate criteria emerge very quickly” (This quotation is taken from Trainer's Guide for the instructional module, “Classroom Assessment Based on Observation and Judgment,” developed by the Center for Classroom Assessment, Northwest Regional Educational Laboratory.)

One technique that may be helpful is to sort examples of actual student work into three piles, the very best, the poorest and those in between. With your colleagues, try to articulate what makes the good assignments good.

In the playground example, an activity that is analogous to viewing examples of student work would be to visit some actual playgrounds. In doing so, the playground rubric developers realized some things should be added to their preliminary list. These include:

- \* equipment for disabled children
- \* good visibility so that parents can see their children at all times
- \* drinking fountain
- \* some shady areas

**Ask yourself these questions:**

- \* What are the attributes of a quality performance of the assessment task?
- \* By what qualities or features will I know whether students have produced an excellent response to my assessment task?
- \* What do I expect to see if this task is done excellently, acceptably, poorly?
- \* Do I have samples or models of student work, from my class or other sources, which exemplify some of the criteria I might use in judging this task?
  - Adapted from Herman, Aschbacher, and Winters (1992)

**STEP 3: REFINE AND CONSOLIDATE YOUR LIST OF DIMENSIONS AS NEEDED.**

Your list will probably contain many more dimensions than you will actually be able to evaluate for each of your students. Try to cluster your tentative list of dimensions into just a few categories or scales. Alternatively, you may wish to develop a single, holistic scale. There is no “right” number of dimensions, but there should be no more than you can reasonably expect to rate. The dimensions you use should be related to the learning outcome(s) that you are assessing.

Returning to the playground example, some possible dimensions or scales might be (1) safety and comfort, (2) appearance, and (3) amount and variety of equipment. Note that there are usually several possible sets of categories that would work. For example, some people might prefer to have separate scales for safety and comfort or they might prefer to group comfort with appearance.

**STEP 4. WRITE A DEFINITION OF EACH OF THE DIMENSIONS.**

You may use your brainstormed list to describe exactly what each dimension encompasses.

In the playground example, “appearance” may be defined to include whether the playground is clean and attractive. This might include being free of graffiti and garbage, surfaces are clean and unchipped, equipment is in good repair, and grass and plantings are well maintained. “Safety and comfort” may be defined as the extent to which the equipment and surroundings are safe and comfortable for children and their parents. This scale could include such features as soft surfaces beneath playground equipment, equipment that does not endanger children, and places for parents to sit that allow them to see their children at all times. “Amount and variety of equipment” could include the types of equipment available, whether equipment is appropriate for different ages, whether there is sufficient equipment and whether there is equipment that can be used by disabled children.

**STEP 5. DEVELOP A CONTINUUM (SCALE) FOR DESCRIBING THE RANGE OF PRODUCTS/PERFORMANCES ON EACH OF THE DIMENSIONS.**

**(A) Describe in words a product/performance that is outstanding.**

For each of your dimensions, what characterizes the best possible performance of the task? This description will serve as the anchor for each of the dimensions by defining the highest score point on your rating scale. Using actual examples of student work as a guide will make this process much easier.

Returning to the playground example, a playground that is exemplary in the safety and comfort category would have cushioned rubber surfaces beneath playground equipment, no splinters or sharp edges, sturdy equipment, a drinking fountain that works, some shady areas, separation of equipment for older and younger children, and places for parents to sit that allow them to see their children at all times.

**(B) Describe in words the worst possible product/performance.**

This will serve as a description of the lowest point on your rating scale. For example, the *least* safe and comfortable playground possible would have concrete beneath the playground equipment, splinters, sharp edges, flimsy or broken equipment, equipment for older and younger children not separated, no shaded areas, no drinking fountain or one that is not working, either few or no places for parents to sit or places that do not afford them an unobscured view of their children.

**(C) Describe characteristics of products/performances that fall at the intermediate points of the rating scale for each dimension.**

Often these points will include some major or minor flaws that prevent the product/performance from receiving a higher rating.

The middle point(s) on the playground safety/comfort scale would describe playgrounds that did not meet all the criteria for the highest scale point (the ideal), but that were better than those described by the lowest scale point.

For example, if 4 were the maximum score on the safety/comfort scale, playgrounds meeting all the criteria of exemplary playgrounds in part (a) would receive a score of 4.

A score of 3 might be assigned to a playground that has cushioned rubber or wood chips beneath the playground equipment, sturdy equipment, no splinters or sharp edges, and at least some seating that affords parents an unobstructed view of their children. The playground may lack shade or a working water fountain. Equipment for older and younger children may not be separated.

A score of 2 might be assigned to a playground that has cushioned rubber or wood chips beneath the playground equipment, some equipment that is not sturdy, some splinters or sharp edges. Seating for parents is inadequate or does not afford them a clear view of their children. The playground lacks shade, a fountain or separation of equipment for older and younger children.

A score of 1, the lowest possible score, would be assigned to playgrounds that fit the description of the worst possible playground in part (b). It would also be possible to define a scale with more or fewer intermediate points.

**How many points should a rating scale have?**

There is no one right answer to this question. Consider these as you make your decision:

- Each point on the scale needs to be well defined. This may be difficult to do for large scales.
- Larger scales make it harder to get good agreement among scorers (inter-rater reliability).
- Extremely short scales make it difficult to identify small differences between students.
- Do you simply want to divide students into two or three groups, based on whether they have attained or exceeded the standard for an outcome? If so, then a short scale may be adequate.
- If you are rating a product/performance on several different dimensions, will you want to add up the scores so that each is equally weighted? If so, you may find it easier to use scales that are all the same length.

**STEP 6. ALTERNATIVELY, INSTEAD OF A SET OF RATING SCALES, YOU MAY CHOOSE TO DEVELOP A *HOLISTIC* SCALE OR A CHECKLIST ON WHICH YOU WILL RECORD THE PRESENCE OR ABSENCE OF THE ATTRIBUTES OF A QUALITY PRODUCT/PERFORMANCE.**

For example, a holistic playground rating scale could have its highest possible score assigned to a playground that is attractive, clean, well-maintained, safe and comfortable for parents and children, has a wide variety of equipment and a sufficient amount of equipment.

A checklist for playground safety and comfort might look something like this:

Mark a check next to each item that describes the playground you are rating.

- soft surface beneath equipment
- no splinters
- no sharp edges
- equipment appears to be sturdy
- equipment in good repair
- shade available
- adequate seating for parents
- parents have a clear view of children
- separate areas for younger and older children
- drinking fountain that works

A total score can be obtained by adding the number of checks.

#### ANALYTICAL AND HOLISTIC RUBRICS

A rubric with two or more separate scales is called an **analytical** rubric. This contrasts with a scoring rubric that uses only a single scale that yields a global or **holistic** rating. ISBE's scoring rubric is actually a hybrid of the two types, since raters also assign an overall, or integration score to each composition.

Holistic scoring is often more efficient, but analytical scoring systems generally provide more detailed information that may be useful in planning and improving instruction and communicating with students.

An example from writing contrasts holistic vs. analytical scores

- When a piece of writing is scored holistically, we give it one overall score.
- All the performance criteria for writing are considered together and the judgments are synthesized into one overall score.

When a piece of writing is scored analytically as it is in our Hawaii State Assessment, information about a student's performance on each of the following criteria is given: meaning, voice, clarity, design, and conventions.

As you think about these two types of rubrics, consider:

- Which would provide the teacher with better diagnostic information to use in planning instruction.
- Which would provide the student with the clearest feedback about his or her work and how to make it better?
- Which would probably be more time-consuming to use?

#### GENERAL VS. SPECIFIC RUBRICS

Scoring rubrics may be specific to a particular assignment or they may be general enough to apply to many different assignments. Usually the more general rubrics prove to be most useful, since they

eliminate the need for constant adaptation to particular assignments and because they provide an enduring vision of quality work that can guide both students and teachers. Nearly all the sample rubrics that appear in this book are general rubrics that may be applied to a wide range of assessments.

#### **STEP 7: EVALUATE YOUR RUBRIC USING THE FOLLOWING CRITERIA.**

##### **Things to consider when selecting a rubric**

- Does the rubric relate to the outcome(s) being measured? Does it address anything extraneous?
- Does the rubric cover important dimensions of student performance?
- Do the criteria reflect current conceptions of “excellence” in the field?
- Are the categories or scales well-defined?
- Is there a clear basis for assigning scores at each scale point?
- Can the rubric be applied consistently by different scorers?
- Can the rubric be understood by students and parents?
- Is the rubric developmentally appropriate?
- Can the rubric be applied to a variety of tasks?
- Is the rubric fair and free from bias?
- Is the rubric useful, feasible, manageable and practical?

#### **STEP 8. PILOT TEST YOUR RUBRIC OR CHECKLIST ON ACTUAL SAMPLES OF STUDENT WORK.**

Your purpose is to see if (1) the rubric is practical to use and (2) whether you and your colleagues can generally agree on what scores you would assign to a given piece of student work. When you pilot test a rubric, it is best to use examples of work that span the entire continuum from very poor to very good.

In the playground example, we would try to use our rubric to rate some actual playgrounds of varying quality.

#### **STEP 9. REVISE THE RUBRIC AND TRY IT OUT AGAIN.**

It is unusual to get everything right the first time, so be prepared to revise your rubric. The pilot test and the criteria shown below can serve as your guides. Did the scale have too many points? Too few? Were the definitions of the score points explicit enough? See evaluating rubrics for more information.

#### **STEP 10. SHARE THE RUBRIC WITH YOUR STUDENTS AND THEIR PARENTS.**

This may entail rewording your rubric in less-technical language, especially for younger students. Give your students a clear target to aim for by letting your students know what constitutes quality work. Judith Arter (Designing Scoring Rubrics for Performance Assessments: The Heart of the Matter, paper presented at the Annual Meeting of the American Educational Research Association, Atlanta, Georgia, April 1993) writes that “assessments communicate and model what we value.” Training students to use the rubric to score their own work can be a powerful instructional tool. Sharing the rubric with parents will help them understand what you expect from their children.