DOCUMENT RESUME

ED 480 070 CG 032 643

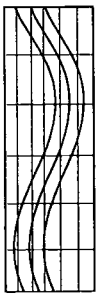
AUTHOR	Perlman, Carole C.
TITLE	Performance Assessment: Designing Appropriate Performance Tasks and Scoring Rubrics.
PUB DATE	2003-08-00
NOTE	12p.; In: Measuring Up: Assessment Issues for Teachers, Counselors, and Administrators; see CG 032 608.
PUB TYPE	Information Analyses (070)
EDRS PRICE	EDRS Price MF01/PC01 Plus Postage.
DESCRIPTORS	Ability; Critical Thinking; Educational Assessment; *Performance Based Assessment; Problem Solving; *Scoring Rubrics; Standards; *Test Construction; *Test Selection

ABSTRACT

Performance assessments can provide an effective means of measuring abilities that are difficult or impossible to measure with a multiple-choice test, such as ability to communicate, solve problems, and employ critical thinking skills. Performance assessments consist of a task and a set of scoring guidelines, or a rubric. Both performance tasks and rubrics must be chosen carefully. This chapter reviews the design of appropriate performance tasks and rubrics. It concludes that a good assessment task is aligned with the standards being measures, requires the student to exercise critical thinking skills, is fair, and is a worthwhile use of instructional time. A well-defined scoring rubric is essential for reliable measurement and to provide students with a clear vision of what constitutes excellent work. (GCP)

Performance Assessment: Designing Appropriate Performance Tasks and Scoring Rubrics

By
Carole C. Perlman



Chapter 35
Performance Assessment
Designing Appropriate Performance Tasks and Scoring Rubrics

Carole L. Perlman

A performance assessment consists of two parts, a task and a set of scoring criteria or a scoring rubric. Unlike a multiple-choice or true-false test in which a student is asked to choose one of the responses provided, a performance assessment requires a student to generate his or her own response. For example, a performance assessment in writing would require a student actually to write something, rather than simply to answer some multiple-choice questions about grammar or punctuation. The assessment task may be a product, performance, or extended written response, ideally one that requires the student to employ critical thinking skills. Some examples of performance-assessment tasks are oral presentations, essays, works of art, science fair projects, research projects, musical performances, open-ended math problems, and analyses or interpretations of literature. Performance assessments are well suited for measuring complex learning outcomes such as critical thinking, communication, and problem-solving skills that may not lend themselves well to a multiple-choice or other forced-response format.

Because a performance assessment does not have an answer key of the type that a multiple-choice test does, scoring a performance assessment necessarily involves making some subjective judgments about the quality of a student's work. A good set of scoring guidelines or rubrics provides a way to make fair and sound judgments by setting forth a uniform set of precisely defined criteria or guidelines for judging student work.

Selecting Tasks for Performance Assessments

The best performance-assessment tasks are interesting, worthwhile activities that relate to your instructional outcomes and allow your students to demonstrate what they know and can do. Some ideas for

performance-assessment tasks in a variety of subjects can be found at http://intranet.cps.k12.il.us/Assessments/Ideas_and_Rubrics/ideas_and_rubrics.html. A very good resource for science performance assessments is Performance Assessment Links in Science (http://www.pals.sri.com). Two excellent sources of information on developing and using performance assessments are Stiggins (1997) and Arter and McTighe (2001). The ERIC Clearinghouse on Assessment and Evaluation at http://www.ericae.net has links to many publications on performance assessment.

As you decide what tasks to use, consider the following criteria, which I have adapted from Herman, Aschbacher, and Winters (1992):

Does the task truly match the outcomes or standards you are trying to measure? This is a must. The task should not require knowledge and skills that are irrelevant to the outcome. For example, if you are trying to measure speaking skills, asking the students to summarize orally a difficult science article would penalize those students who are poor readers or who lack the scientific background to understand the article. In that case, you would not know whether you were measuring speaking or (in this case) extraneous reading and science skills. Sometimes it is possible to enable students to perform successfully despite gaps in prior factual knowledge by giving them access to textbooks or reference materials.

Does the task require the students to use critical thinking skills? Is recall all that the task assesses, or must the student analyze, draw inferences or conclusions, critically evaluate, synthesize, create, or compare? In general, when you are assigning a performance task, students should not have received specific instruction in its solution. If students know the solution you may be measuring simply rote memory. For example, suppose an instructional outcome deals with analyzing an author's point of view, and you devote a class discussion to an analysis of the authors' points of view in two editorials. If you then ask the students to write an essay analyzing the authors' positions in those editorials, you are essentially measuring students' recall of the class discussion, rather than their ability to analyze point of view. A better assessment would be to ask the students to analyze editorials that have not been discussed in class, in order to see how well they can generalize their knowledge and skills to a novel piece of writing.

Is the task a worthwhile use of instructional time? Performance assessments may be time-consuming, so it stands to reason that that the time should be well spent. Instead of being an add-on to regular instruction, the assessment should be part of it.

Does the assessment use engaging tasks with real-world application? The task should capture students' interest enough to ensure that they are willing to try their best. Does the task represent something important that students will need to do in school and in the future? Many students are more motivated to do a task when they see that it has some meaning or connection to life outside the classroom.

Are the tasks fair and free from bias? Is the task an equally good measure for students of different genders, races, cultures, and socioeconomic groups represented in your school population? Will all students have equivalent resources—at home or at school—with which to complete the task? Have all students received equal opportunity to learn what is being measured?

Is the task clearly defined? Are the instructions for teachers and students clear? Do students know exactly what is expected of them?

Is the task feasible? Can students reasonably be expected to complete the task successfully? Will you and your students have enough time, space, materials, and other resources? Does the task require knowledge and skills that you have taught or are able to teach?

Will the task be credible? Will students, parents, and your colleagues view the task as being a meaningful, challenging, and appropriate measure?

Understanding Scoring Rubrics

A scoring rubric has several components, each of which contributes to its usefulness. These components include the following:

- one or more dimensions on which performance is rated
- definitions and examples that illustrate the attribute or attributes being measured
- a rating scale for each dimension

Ideally, the rubric should be accompanied by examples of student work that illustrate each level of the rating scale. The rubric should

organize and clarify the scoring criteria well enough that two raters who apply the rubric to a student's work will generally arrive at the same score. The degree of agreement between the scores assigned by two independent scorers is a measure of the reliability of an assessment. This type of consistency is especially important if assessment results are to be aggregated across classrooms, schools, or districts.

Analytical Versus Holistic Rating

A rubric with two or more separate scales—for example, a science lab rubric divided into sections related to hypothesis, procedures, results, and conclusion—is called an *analytical rubric.* A scoring rubric that uses only a single scale yields a global or *holistic,* rating. In a holistic rating system, the overall quality of a student's response might, for example, be judged excellent, proficient, marginal, or unsatisfactory. Holistic scoring is often more efficient, but analytical scoring systems generally provide more detailed information that may be more useful in planning and improving instruction and communicating with students.

Whether you choose and analytical or holistic rubric, you must clearly label and define each point on the scale. There is no best number of scale points, although it is generally advisable to avoid scales with more than six or seven points. With very long scales, it is often difficult to differentiate adequately between adjacent points (e.g., on a 100-point scale, it would be difficult to explain why you assigned a score of 81 rather than 80 or 82). Different scorers are also less likely to agree on ratings when very long scales are used. Extremely short scales, on the other hand, make it difficult to identify small differences between students. A short scale may be adequate for some purposes, however, such as when you simply want to divide students into two or three groups, based on whether they have failed to attain, attained, or exceeded the standard for an outcome.

A good rule of thumb is to have as many scale points as can be well defined and can adequately cover the range from very poor to excellent performance. If you decide to use an analytical rubric, you may wish to add or average the scores from each scale to get a total score. If you feel that some scales are more important than others (and assuming that the scales are of equal length), you may give them more weight by multiplying those scores by a number greater than one. For example, in the case of a science lab write-up, if you felt that the results section scale was twice as important as all the others, you would multiply the score on that scale by two before you added up the scale scores to get a total score.

Specific Versus General Rubrics

Scoring rubrics may be specific to a particular assignment or may be general enough to apply to many different assignments. Usually general rubrics prove to be more useful, because they need not be constantly adapted to particular assignments and they provide an enduring vision of quality work that can guide both students and teachers. Some states and districts have adopted a set of standard scoring rubrics; in that case, it is advisable to use those rubrics for classroom assessments whenever possible to avoid the potential for confusion when two or more different rubrics are used to score similar assignments.

A rubric can be a powerful communication tool. When shared among teachers, students, and parents, the rubric informs everyone about what characteristics of student work are most highly valued. It provides a means for you and your colleagues to clarify your vision of excellence and convey that vision to your students and their parents. It can also provide a rationale for assigning grades to subjectively scored assessments. Sharing the rubric with students is only fair and is necessary if we expect them to do their best possible work. An additional benefit of sharing the rubric is that students are empowered to critically evaluate their own work.

In order for a rubric to be effective in communicating what we expect of our students, students and parents must be able to understand it. This may require restating all or part of the rubric to eliminate educational jargon and explain the criteria in a way that is appropriate for the students' developmental level. (For example, "The story has a beginning, middle, and end" is clearer and more helpful to students than "Observes story structure conventions.")

Selecting a Scoring Rubric

Teachers interested in using rubrics to assess performance-based tasks have three options: use an existing rubric as is, adapt or combine rubrics to suit a specific purpose, or create a rubric from scratch. One online source of rubrics is the Chicago Public Schools' rubrics bank (Perlman, 1994) at http://intranet.cps.k12.il.us/Assessments/Ideas_and_Rubrics/Rubric_Bank/rubric_bank.html. Some state departments of education have rubrics and scored examples of student work available on their websites. Links to state education agencies may be found at the Council of Chief State School Officers website: http://www.ccsso.org.

If you are considering using an existing rubric, ask yourself these questions:

- 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 different scorers apply the rubric consistently?
- Can students and parents understand the rubric?
- Is the rubric developmentally appropriate?
- Is the rubric applicable to a variety of tasks?
- Is the rubric fair and free from bias?
- Is the rubric useful, feasible, manageable, and practical?

In order to have an existing rubric better suit your task and objectives, you might make the following adaptations:

- Reword parts of the rubric.
- Drop or change one or more scales of an analytical rubric.
- Omit criteria that are irrelevant to the outcome you are measuring.
- Mix and match scales from different rubrics.
- Change the rubric for use at a different grade level.
- Add a "no response" category at the bottom of the scale.
- Divide a holistic rubric into several scales.

If adopting or adapting an existing rubric does not work for your purposes, here are some steps to follow in developing your own scoring rubric:

1. With your colleagues, make a preliminary decision on the dimensions of the performance or product to be assessed. The dimensions you choose may be guided by national curriculum frameworks, publications of professional organizations, sample scoring rubrics (if available), or experts in the relevant subject area. Alternatively, you and your colleagues may brainstorm a list of as many key attributes of the product or performance to be rated as you can. In brainstorming, consider what you look for when you grade assignments of this nature and which elements of this product or performance you emphasize during teaching.

2. Look at some actual examples of student work to see if you have omitted any important dimensions. Try sorting examples of actual student work into three piles: the best, the poorest, and those in between. With your colleagues try to articulate what makes the good assignments good.
3. Refine and consolidate your list of attributes as needed. Try to cluster your tentative list of dimensions into a few categories or scales. Alternatively, you may wish to develop a single, holistic scale. There is no absolute number of dimensions you should generate, but there should be no more than you can reasonably expect to rate. The dimensions you use also should be related to the learning outcomes you are assessing.
4. Write a definition of each dimension. You may use your brainstormed list to describe exactly what each dimension encompasses.
5. Develop a continuum (i.e., scale) for describing the range of products or performances on each dimension. Using actual examples of student work to guide you will make this process much easier. For each dimension, ask yourself what characterizes the best possible performance of the task. This description will serve as the anchor for that dimension by defining the highest score point on your rating scale. Next describe in words the worst possible product or performance. This will serve as a description of the lowest point on your rating scale. Then describe characteristics of products or performances that fall at intermediate points of the rating scale for each dimension. Often these points will describe some major or minor flaws that preclude a higher rating.
6. Alternatively, instead of generating a set of rating scales, you may choose to develop a holistic scale or a checklist on which you can record the presence or absence of the attributes of a high- quality product or performance.
7. Evaluate your rubric using the questions listed previously.
8. Pilot test your rubric or checklist on actual samples of student work to see whether it is practical to use and whether you and your colleagues generally agree on what scores you would assign to a given piece of work.
9. Revise the rubric and pilot test it again. It is unusual to generate a perfect the first time. Ask yourself these

questions: Did the scale have too many or too few points? How could the definitions of the score points be made more explicit?

10. Share the final rubric with your students and their parents. 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 and clarify what constitutes excellent work.

Some Considerations in Using Performance Assessments

Performance assessments have advantages and disadvantages. On the plus side, they can provide rich learning experiences; they can simulate real-world problem solving; they can encourage students to critically evaluate their own work; they can provide teachers with insights into their students' cognitive processes; they can foster good instruction; and they can be an excellent measure of students' abilities to synthesize, evaluate, and solve problems. Learning to use a scoring rubric can be an excellent staff development experience for teachers. Finally, some instructional outcomes simply do not lend themselves well to other assessment formats. What are the downsides? Performance assessments can be expensive and time-consuming to administer and score, particularly when they are part of districtwide or statewide assessment. Assessment results are generalizable to the extent available evidence shows that scores on one assessment predict how well students perform on another assessment of the same outcome; a good result on one performance task may not generalize well to similar tasks. The subjectivity inherent in scoring a performance assessment may make some people uncomfortable, although a well-constructed rubric coupled with effective rater training and monitoring can go a long way toward addressing those concerns. Finally, certain kinds of knowledge and skills are more efficiently assessed using other assessment formats, such as multiple-choice tests.

An assessment is reliable if it yields results that are accurate and stable. In order for a performance assessment to be reliable, it must be administered and scored in a consistent way for all students who take the assessment. Once you decide on a rubric, the best way to promote reliable scoring is to have well-trained scorers who thoroughly understand the rubric and who periodically all score the same samples of student work to ensure that they are maintaining consistent scoring.

Another way to increase reliability is to adhere carefully to the rubric as you score student work. Not only will this increase reliability and validity, but it is only fair that the agreed-upon rubric that you have shared with students and parents is what you actually use to rate student work. Nonetheless, human beings making subjective judgments may unintentionally rate students based on things that are not in the rubric at all. Therefore, the conscientious scorer will frequently monitor his or her thinking to prevent extraneous factors from creeping into the assessment process.

Summary

Performance assessments can provide an effective means of measuring abilities that are difficult or impossible to measure with a multiple-choice test, such as ability to communicate, solve problems, and employ critical thinking skills. Performance assessments consist of a task—for example, a project, extended written response, oral presentation—and a set of scoring guidelines, or a rubric. Both performance tasks and rubrics must be chosen carefully. A good assessment task is aligned with the standards being measured, requires the student to exercise critical thinking skills, is fair, and is a worthwhile use of instructional time. A well-defined scoring rubric is essential for reliable measurement and to provide students with a clear vision of what constitutes excellent work. Educators may design their own performance-assessment tasks and rubrics, or they may use or adapt tasks and rubrics created by their state or district educational systems. The Internet is a good source of sample performance-assessment tasks and rubrics.

Portions of this chapter were adapted from C. L. Perlman (2002), An introduction to performance assessment scoring rubrics, in C. Boston (Ed.), *Understanding scoring rubrics: A guide for teachers,* College Park, MD: ERIC Clearinghouse on Assessment and Evaluation, and from C. L. Perlman (1994), *The CPS performance assessment idea book,* Chicago, IL: Chicago Public Schools.

References

Arter, J., & McTighe, J. (2001). *Scoring rubrics in the classroom: Using performance criteria for assessing and improving student performance.* Thousand Oaks, CA: Corwin.

✦Herman, J., Aschbacher, P., & Winters, L. (1992). *A practical guide to alternative assessment.* Alexandria, VA: Association for Supervision and Curriculum Development.

Perlman, C. L. (1994). *The CPS performance assessment idea book.* Chicago, IL: Chicago Public Schools.

Stiggins, R. J. (2001). *Student-centered classroom assessment* (3rd ed.). Upper Saddle River, NJ: Merrill.

✦ Document is included in the Anthology of Assessment Resources CD

U.S. Department of Education
Office of Educational Research and Improvement (OERI)
National Library of Education (NLE)
Educational Resources Information Center (ERIC)

NOTICE

Reproduction Basis

	This document is covered by a signed "Reproduction Release (Blanket)" form (on file within the ERIC system), encompassing all or classes of documents from its source organization and, therefore, does not require a "Specific Document" Release form.
X	This document is Federally-funded, or carries its own permission to reproduce, or is otherwise in the public domain and, therefore, may be reproduced by ERIC without a signed Reproduction Release form (either "Specific Document" or "Blanket").