

Evaluating Classroom Learning

Introduction

As you work with your colleagues to design an arts-integrated unit using the Perpich Arts Integration Network of Teachers Planning Guide you will want to establish clear criteria for determining the quality of learning demonstrated by student's successful completion of the assessment activity. Clear criteria for evaluation that are aligned with the kind of learning called for in the benchmark and learning goal help students understand what good work looks like, and to determine the next steps they need to take in the learning process. Once teachers have information from the assessment activity by watching, reading, or listening to the student work, they are able to evaluate the quality of the learning. **Evaluation is the judgment about the level of achievement in student learning.**

Evaluative Criteria

If the assessment activity provides the direct evidence of student learning then the teacher's application of evaluative criteria is her judgement about the quality of learning that can be read in the evidence.

Pencil and paper tests are interpreted by teachers using the **scoring guide**, a set of the correct answers. Closely aligned to the scoring guide is the **checklist**, a way to record the presence or absence of a student's accurate response. He either did it, or didn't do it.

You may have used the x + - system to evaluate work that is proficient, exceeds proficiency, or is below proficiency. This is a **rating scale**. These big clusters of three or more levels, or ratings, of performance may be easy to apply, yet invite imprecision because their application is influenced by criteria that can be different at each level. They can however, be helpful in generating descriptions of quality as teachers work to distinguish each level.

With the advantage of organizing consistent criteria that develop equally over three or more levels of performance, the **rubric** is the best tool for working to evaluate the quality of substantial assessment activities. Remember though that not every assessment activity warrants a rubric. Our goal over time is that assessment and evaluation is **multiple** (includes more than one measure) and **varied** (applied activities beyond a pencil and paper test), **efficient** (involves teachers' time and effort that is proportional to its instructional value) and **effective** (aligns well to do the job teachers and students need it to accomplish).

Tips for Designing Effective Evaluative Criteria

1. Make sure that the evaluative tool (rubric or rating scale) includes the most important traits, given the purpose of the assessment and the qualities of excellent performance. Consider the following:

- Are you evaluating what is easy to score rather than what is most important?

- Could a student meet all the evaluative criteria and get high score without really demonstrating the desired understanding or producing excellent work?
- Are any of the criteria or reasons for the evaluation arbitrary? In other words, are you giving or taking away points based on characteristics that have little to do with meeting the benchmark/learning goal?

2. Beware of the following common problems with evaluative tools:

- Evaluating the length of the paper instead of its quality.
- Focusing on mechanics, organization, and presentation rather than content, substance, and effect. For example, a science project display could be attractive but superficial.
- Looking for quantity rather than results (e.g., the number of information sources used in research instead of the appropriateness and thoroughness of those sources; number of reasons in a persuasive essay instead of the logic of the reasoning).
- Demanding that the performance follow an arbitrary format (e.g., five-paragraph essay), even though expert performance follows different forms or approaches.

3. Check for consistency of the descriptive terms throughout the evaluative scale. For example, if the top score point includes the descriptors consistently and thorough, we would expect to see parallel descriptors in the lower score points—for example, sometimes and incomplete.

4. Use the following prompts to help avoid these problems:

- Since the aim of [the performance] is to provide evidence of _____ , we need to assess whether or not the performance has been _____ .
(List appropriate traits, given the purpose of the task and the evidence it demands.)

- The best pieces of work are those that are always _____
(insert traits from your list) Does the sentence make sense or not? If yes, the trait is appropriate; if not, it is probably arbitrary.

Source: *The Understanding by Design Guide to Advanced Concepts in Creating and Reviewing Units*. Grant Wiggins and Jay McTighe (2012) ASCD Alexandria, Virginia.