

How Many Levels of Quality Should We Represent in a Rubric?

Horror vacui. *Aristotle*

Aristotle may or may not be right when he said *nature abhors a vacuum*, but when we create a rubric, academics loathe leaving cells empty.

Problems arise when a rubric contains many quality levels, creating empty cells that demand descriptive language:

- We have difficulty writing unambiguous descriptions of quality for rubrics with 4 or more levels. Rubrics with many levels introduce fine-grain distinctions that we cannot describe clearly. With ambiguous definitions, we have difficulty making consistent decisions.
- Some characteristics of student work are best described and evaluated with 2 or 3 levels (thesis statement included / not included; problem solved accurately / problem solved with minor technical errors / problem not solved).
- Complex rubrics create the risk that reviewers ignore the language used to define levels of quality and instead use unwritten assumptions when they evaluate work. A rubric based on five levels of quality is especially vulnerable; reviewers might adopt a mental shortcut in which 5 represents A-level work, 4 represents B-level work, etc. Other reviewers might rely on the descriptive language to make decisions. If both reviewers are grading the same work, the reviewers will disagree and the first reviewer might use the rubric inconsistently.

Humphry and Heldsinger (2014) reported that reviewers' judgments can be influenced by the structure of the rubric. A typical rubric creates the same number of quality levels for every criterion element in the rubric. When using these rubrics, reviewers tended to make global judgments about student papers and assign similar ratings for each element in the rubric. Thus, if a student scores in the top category on one element, the student is likely to receive scores in the top category on all other elements, even when performance across elements is uneven.

In contrast, reviewers who make decisions based on the element descriptions are more likely to assign different ratings for different elements, revealing patterns of strength and weakness in student performance. A rubric that uses different numbers of quality levels for different criteria focuses reviewer attention on the descriptive language and improves consistency. For example, a rubric might use 5 levels of quality to describe the use of evidence to build and support an argument but use only 3 levels of quality to describe the use of mechanics of language (significant errors that interfere with readability or communication, occasional errors that do not interfere with readability or communication, no errors or minor errors, mostly related to lapses in proofreading). Reviewers who used rubrics with varying numbers of quality for different criteria

were more likely to make independent decisions about each criterion (Humphry & Heldsinger, 2014).

Consider the merits of a rubric that (gasp) has some empty cells and uses 4 or 5 levels of quality to evaluate some components of student work and uses 2 or 3 levels of quality to evaluate other components. If the empty cells bother you or your students, fill them with a grey shading to let everyone know you intended them to be blank. Nilson (2014) describes rubrics with only 2 levels in which each element aligns with competency on a specific student learning outcome.

Whether you use the same number of levels for all elements in a rubric or allow these to vary, pay attention to how you combine rubric elements to generate the final grade for the assignment. Make sure that scores on the most important elements contribute most to the final grade. Rubrics that include many elements (even 2 or 3 level elements) for minor details that are easy to evaluate can come to dominate the total score. If you want the content of a literature review paper to determine 30% of the grade, ensure that it contributes 30% of the possible points to the final score. Use multipliers for a rubric element to increase the contribution of critical elements to the final score.

If the rubric generates data for assessment purposes, keep a record of the scores students earn on each element. These individual element scores are critical for identifying areas of strength and weakness in student work. For example, students may earn high scores on an element that evaluates how accurately students describe content of the literature they review, but they may earn lower scores on how well they interpret and apply that content to analysing a problem or building an argument.

Resources

Humphry, S. M., & Heldsinger, S. A. (2014). Common structural design features of rubrics may represent a threat to validity. *Educational Researcher*, 43, 253-263.
doi:10.3102/0013189X14542154

Nilson, L. B. (2014). *Specifications grading*. Sterling, VA: Stylus.

Submitted by:

Claudia J. Stanny, Ph.D., Director
Center for University Teaching, Learning, and Assessment
University of West Florida
Pensacola, FL
<http://uwf.edu/offices/cutla/>