

Enhancing the Reliability of Rubrics

Even the most carefully constructed rubric may be used in ways that unintentionally result in scoring errors and biases. Errors may include:

Leniency errors: a faculty member judges student work better than most other colleagues.

Errors of scale: Generosity (using only the high end of the rating scale), Severity (using only the low end of the rating scale), Central Tendency (avoiding both ends of the scale).

Halo and Contamination effects: faculty know the particular student and give higher scores based on rapport or lower scores based on dissatisfaction with the student; irrelevant factors such as handwriting or ethnic bias impact scoring. .

First impression bias: A poorly constructed essay introduction or a stained portfolio folder may distort an overall judgment.

Contrast effect bias: Students are compared against other students instead of established standards. The weakest paper in a group may be assigned the lowest rating on a scale even though the paper meets minimally acceptable standards.

Rater drift: When tired of assessing large numbers of an assignment, some faculty become more stringent while others aim to finish quickly and score more leniently.

Strategies for minimizing errors and biases typically include the following:

- 1) Remove identifying student information.
- 2) Meet with other faculty to discuss each performance level. Then score a few samples of student work, discussing and resolving any differences in ratings. Score another sample of student work. When you are reasonably sure that you are interpreting the rubric consistently, begin actual scoring. Typically, two-three cycles results in calibration. Such sessions may need to be held at regular intervals at first.
- 3) Each sample of student work is read by two faculty. If they disagree, a third breaks the tie.
- 4) After scoring many samples of student work, rescore a few early samples to guard against rater drift.

How seriously should we take potential problems with bias?

If scores are reasonably consistent across faculty, scoring typically proceeds. However, if faculty in a department don't feel you can use results with confidence, you will want to invest some time in discussions of ratings (#2 above).

If a particular assignment is only one of many measures of student learning, rigorous efforts to eliminate scoring biases may not be worth the investment of time and resources. If a particular assignment is going to be used to make major decisions about a program or curriculum, ensuring accurate and consistent reporting is important.