

QUICK GUIDE TO NORMING ON STUDENT WORK FOR PROGRAM-LEVEL ASSESSMENT

Office of Assessment of Teaching and Learning, Washington State University, June 2016

Norming, or calibration, is a process that brings a group of faculty raters together to decide how to assess student work in a consistent way, so that regardless of which rater assesses the work, the rating falls within a close range. The process is generally overseen by a facilitator, either from within the department or from outside, someone familiar with norming processes, using a rubric or similar scoring tool.

This quick guide offer a general approach for norming faculty who will be scoring student work for program-level assessment of student learning. Because each department, program and situation is different, we recommend that you consult with ATL to develop a process that is best for your specific circumstances, including the kind of student work you're assessing (such as, paper, project, presentation or performance) and the team of raters.

Norming for Program-level Assessment

Assessment at the program-level differs from grading student work in the course. Grading is the process of evaluating how well a student completed a given assessment or given class. Grading may include issues such as effort, timeliness, how much the student has improved in the semester, none of which necessarily reflect how students perform on specific learning outcomes. In program-level assessment, raters must recognize that they are not grading individual students, but rather providing information to the program about what and how well students are learning. (Instructors might also use a similar norming process to develop more consistent grading in a particular course.)

Because the purpose of program-level assessment is to determine how well the students are achieving the student learning outcomes, the students' class standings do not matter. For example, if raters are assessing first year student papers, they will not give additional points because a student's work is *good considering they are just starting at the university* or because it *would deserve an A in an introductory class*. Rather, faculty rate the student papers at the level the work demonstrates. The same holds true for student work from all class levels. While most programs expect their first year students to perform at the lower end of the rubric, and sophomores a bit higher, etc., raters should not, as a matter of course, assume what student scores will be before they rate the student work or assign points for anything except what is demonstrated in the student work. In this way, program-level assessment will generate data about student learning across the curriculum.

The Norming Process

Faculty raters come together to

- a. Practice using the rubric on several samples of student work
- b. Discuss scores and develop a shared understanding of how to apply the criteria at the program level
- c. Develop consensus on scoring, so that reliable data can be generated for program assessment
- d. Potentially: Identify sample work as "anchors" available to other raters, to provide clear examples of what different scores or levels of performance, look like. (Contact ATL for more information about developing or using anchors.)

The Norming Session

A group of faculty raters and a facilitator meet for 1-2 hours. Ideally, at least two samples of student work will be used. Having a range of student work quality (i.e., high-performance, mid-level performance and low performance) gives faculty members a chance to understand each other's perspectives about what constitutes strong performance. This process should be adjusted if the program has anchor papers available, a good practice as assessment matures.

1. **Context:** Information should be provided about the reason for the norming, how the information will be used and shared, how this session contributes to program assessment, how long the session should take, and what is expected of the participants.
2. **Materials:** Participants should take some time to get to know the rubric. Clarify if the ratings are to be holistic (one score for the entire rubric) or analytic (one score for each element on the rubric). Clarify if the norming session is for one element of the rubric, some elements of the rubric or for all elements rubric. (For example, it may not be necessary to score on all rubric elements if the program's assessment is focusing on one 1-2 learning outcomes. Choosing a couple of learning outcomes is a good way to get started, focus attention, and manage the time investment.)
3. **Allow Time for Individual Ratings:** Participants are given time to read the first piece of student work, keeping the rubric in mind, or, if they were given the work in advance, to review the work and score it using the rubric. During this process, participants are looking at the student work through the lens of the rubric: what language of the rubric best describes the student work? At the same time, participants can note parts of the student work that they see as representative of the rubric language. When finished reading, participants should score the paper on their own (without discussion with other participants).
4. **Discussion and Consensus:** Scores are collected from all raters, and the group looks for patterns, where scores align and where they differ. Participants should be prepared to talk about their rating and why they see the rating as appropriate. They should also be prepared to listen to colleagues' perspectives and be open to those perspectives. Participants talk about how/if they rated the same or different and why. Participants should talk through all questions and concerns. The goal of this conversation is that raters share their perspective in order to come to an understanding so that they can rate student work with a level of consistency among them. During this conversation, it is important to remember that the participants are deciding how they will assess the student work as a group. If individuals don't agree on a rating, they should attempt to meet a middle ground where all participants are confident they can rate in the same way – if they can interpret and apply the rubric descriptions similarly -- even if they would grade differently in their own classes. (See *More about Coming to Consensus* below)
5. **Repeat as Needed:** Ideally, this process is repeated two or three times with a range of samples, so that participants can clearly see how, when and why student work is rated. Allowing participants the chance to rate a low, medium and high paper can give them experience in how to rate specific examples.

More about Coming to Consensus

The goal of the norming session is for participants to rate in a consistent manner for the same or similar reasons. If, after the process, raters are within one point on a six-point scale, the group can be considered normed. For example, if a rubric is on a six-point scale, and all of the participants align, after discussion, that a student paper is either a 3 or a 4, a facilitator can usually consider the group normed.

Rating and Next Steps

Soon after a norming session, raters generally score additional student work to generate data for program assessment. (The norming session itself calibrates the raters but does not generate assessment data.) For example, after the norming session, raters can score additional papers over a two-week period, without needing to meet again. Anchor papers with scores can be made available electronically for raters to refer to.

Programs are welcome to contact ATL for information on collecting student work (including decisions about sample size and representation), preparing student work for rating, collecting scores from raters, using multiple reads to calculate reliability, and data analysis and presentation for subsequent faculty discussion.

Examples

Program A

Program A has a capstone course with final papers that cover many of the student learning outcomes. The program has randomly pulled a representative sample from the class of seniors to be rated. A group of faculty members have been identified to rate the papers. They assemble at the same time to participate in a norming process.

During norming, several questions arise about the nature of the student learning outcomes and how those are taught in the curriculum. The committee reads a total of four papers, chosen to illustrate a range of levels, coming to consensus on the scoring. Afterwards, committee members each read an assigned number of collected papers and submit their scores for further analysis and results. Two papers are scored by all raters, to monitor reliability.

Program B

In Program B, most of the faculty members have been requested to participate. With such a large group, finding a single time to bring all faculty members together to rate has been difficult.

In order to facilitate the process, faculty members have agreed to the following process: Several norming sessions have been established. The first one is of a representative subgroup of faculty members who have been tasked with creating anchor documents. The documents are then taken to the rest of the norming sessions to norm faculty to the scores of the students' work. This process allows the faculty members to have several, smaller norming sessions and to assure that the faculty are all normed to the same scores for the same papers.

Based on Good Practice Resources

[Calibration Protocol for Scoring Student Work](#), Rhode Island Department of Education

[Scoring Rubric Group Orientation and Calibration](#), University of Hawaii Manoa