

STAT 213, SPRING 2018: GRADING SYSTEM

MY PHILOSOPHY ON GRADES AND GRADING

I find that the stress, anxiety and pressure that grading creates can detract from deep engagement with learning, incentivizing gold star-collection and hoop-jumping, and putting many students to the right of their optimal location on the arousal-performance curve (look up the *Yerkes-Dodson law* on Wikipedia, for example), harming their capacity to learn. At the same time, the outside world expects some signal about the quality of work done in a course, and Oberlin College requires grades to be attached to courses. Moreover, I recognize that, realistically, we all tend to prioritize the parts of our work that “count for something”. So, we are stuck with grades.

My overarching goal in structuring assessment, therefore, is to preserve the signaling property of course grades as a measure of how well you have internalized the content and competencies central to the course by the end of the semester, while reducing to the extent possible the “gold star”/“psychological threat” nature of grading throughout the course, focusing on feedback over judgment.

FOUR GOALS OF LEARNING-FOCUSED GRADING

To this end, I use a non-traditional grading system designed around the following principles, which I attempt to implement as described.

- I. **Goal:** Feedback and marks should align with levels of understanding and competency for specific content, rather than being an undifferentiated mass of ill-defined “points”.
 - **Implementation:** Each graded item (think quiz/exam question) is attached to one or more “specific learning objectives” (SLOs), and evaluated on each of those objectives. Grades are first combined within each SLO, and averaged over SLOs at the end. This means you won’t get numeric scores on quizzes or exams; instead you will have a running measure of your understanding of each concept or skill. This should help me and you identify what you haven’t yet mastered, to help focus my and your effort.
- II. **Goal:** There should be as few “choke points” as possible: moments when, if you can’t demonstrate mastery, your course grade is irreparably harmed.
 - **Implementation:** There are numerous opportunities for “reassessment”; that is, to show understanding of content later, even if you didn’t do well on the quiz or exam when that content was initially assessed. However, unlike in most systems with “extra credit” opportunities, previous grades can only be replaced by revisiting the same content.
- III. **Goal:** Two students who understand the material equally well by the end of the semester should get the same grades, even if it took one longer for that understanding to crystalize.
 - **Implementation:** Your grade on a particular SLO at the end will in most cases be no lower than the grade the *last* time it is assessed.
- IV. The standard for grading on individual assignments should be high, so that nearly everyone has room for improvement.
 - **Implementation:** Individual assignment grades will likely be lower than might be expected in a traditional grading system due to the high standard for mastery. This is counterbalanced by the numerous reassessment and grade replacement opportunities.

My hope is that this system fosters a sense of forward momentum, decreasing grade anxiety, and emphasizing understanding over checking off boxes.

CONCEPTS AND CONTENT (60% OF COURSE GRADE)

Specific Learning Objectives. Specific “bite-sized” concepts and competencies for this course are listed in a set of 65 “Specific Learning Objectives” (SLOs) below. Each homework set, quiz and exam question will be attached to one or more of these SLOs and marked separately for each. In addition, each of the two major “data analysis” projects will involve several content SLOs as well as the integration and communication SLOs (see below).

A list of all the SLOs in the course is in the document entitled “Specific Learning Objectives”.

Grading for Concepts and Content. Graded SLOs can appear in two modalities: in-class assessments (quizzes or exams), and projects. For each item, mastery the associated SLOs will be assessed using an M/M-/P/N scale (“Mastery”, “near-Mastery”, “Progress” and “Not Assessable”). These count as 4, 3, 2 and 0, respectively, for the purposes of averaging.

- For SLOs assessed only in class, the course grade is either the average of all item grades attached to that SLO or the highest exam grade (whichever is higher).
- For SLOs assessed only on projects, the SLO grade is just the highest project grade.
- For SLOs assessed in both modalities, the course grade is the average of these two.
- The overall “Concepts and Content” grade is on a 0-4 scale, and is simply the average SLO grade over all 65 SLOs (or however many wind up being assessed, in the likely event that some wind up being omitted).

Although homework questions are attached to SLOs to help you recognize what kinds of questions correspond to what SLOs, homework is only graded for “good faith effort”, and does not contribute to the “Concepts and Content” component of the grade.

Reassessments on Concepts and Content. By default, the course grade on an SLO which is assessed in class is either the average over all items, or the highest exam item grade. So if a quiz grade is low, that grade can be wiped out and replaced by an exam grade later. There is also the opportunity to replace grades by taking “replacement quizzes” during my office hours. These quizzes have the same structure as regular in class quizzes, and the associated SLO grades replace the lowest in class

item grade. They can replace exam grades for the purposes of computing “average item grade”, but count as quizzes when computing the “highest exam grade”.

To do a reassessment, make an appointment by email (if you ask me about it face-to-face I will still ask you to email me, as I use my email inbox to keep track of these things), identifying an office hour time *which is at least 48 hours in the future* when you would like to do it, and identifying which group of SLOs you want to revisit. You may do at most *one quiz worth of SLOs per week*, and the SLOs you choose should align with “top-level” quiz or exam questions (since not every arbitrary set of SLOs make sense together).

APPLICATION, INTEGRATION AND COMMUNICATION (20% OF COURSE GRADE)

This course is not only about learning specific, discrete statistical concepts; it is also about becoming fluent in the application of these concepts in open-ended data analysis scenarios, and communicating one’s findings to a lay audience. There are three main opportunities to engage in this integration, application, and communication of ideas in more open-ended settings, in the form of two required data analysis projects, and one optional final project. The two required projects will involve linear and logistic regression modeling, respectively; the final project may involve either or both of these, as well as analysis of variance.

In addition to assessing your understanding of specific concepts as outlined above, these projects will be graded on the following four dimensions (each on a 0-10 scale):

- (1) Overall technical soundness: To what extent are the tools you apply appropriate to the questions you are trying to answer, and to what extent are the technical details correctly executed in the code?
- (2) Chain of reasoning: How well are the decisions you make throughout your analysis motivated in terms of the research question and in terms of the results obtained so far?
- (3) Interpretation of results in context: How well did you take the results of your analysis and connect it back to the real world context of the problem, in such a way that a reader not trained in statistics can take something away from your analysis?
- (4) Clarity of communication: How easy is it to follow your writeup? This criterion comprises both the quality of the writing itself, the organization of

the report (are text sections, graphs, etc. well placed for the reader to follow what is going on?), and the aesthetic quality of the report (have you suppressed unnecessary/distracting output, are your figures well labeled and visually appealing?)

Calculation of Overall “Applications” Grades. Although the two projects involve different types of models, and thus the conceptual components are different, the big picture components are largely the same, and my hope is that you will learn to improve your statistical communication as you go. As such, your course grade on each of these four components will be the highest of the following computations for each dimension:

- (1) the average of the grades for that dimension on the two required projects
- (2) the grade on the second project alone
- (3) the grade on the optional final project alone

Example. Suppose a student receives scores of 7, 9, 10, and 8 on the four criteria on the first project; 9, 8, 9 and 9 on the second, and 8, 9, 9 and 10 on the final. Their “technical soundness” grade would be 9, based on the second project. Their “reasoning” grade would be 9, based on the final project. Their “interpretation” grade would be 9.5, based on the average of the first two projects, and their “clarity” grade would be 10, based on the final, for a total of 37.5/40.

The principles here are that (a) you need to make a strong effort on the second required project (i.e., you can’t rest on your laurels based on the first one), (b) there is always the option to “wipe the slate clean” going into the final project, (c) the final project can’t drag down the grade you would otherwise have, and, (d) since the four components are calculated separately, you can gain more by focusing on your areas of weakness.

GRADING OF WRITING PROMPTS AND PROBLEM SETS (10% EACH)

I expect consistent engagement with weekly homework and daily writing prompts. I define “consistent engagement” as completion of 90% of the writing prompts (so, you can miss about 4 classes worth before it starts to affect your grade), and making a “good faith effort” on 90% of problem sets (so, you can miss one for free). If you aren’t able to finish the reading or you missed class on a particular day, just say so; don’t try to B.S. a response. On problem sets, it is better to do a careful job on most of the problems and miss one or two entirely than to turn in a rushed effort in order to attempt every problem.