

Student Self-Assessment Case Study Report

Jan 25, 2012

Aaron Rendahl, . . .

Recap of Situation In this case, the client wants to test if student self-assessment is “as good as” faculty assessment for demonstrating that students have mastered certain skills. They have data on both student and faculty grades, on a five point A–F scale, for 60 students in 4 categories. Student participation was optional; about 60% of students did participate. Finally, faculty evaluation was done by six faculty members, who each evaluated ten students, so that each student was evaluated by only one faculty member.

Possible Issues with Study We are concerned that the students who participated are not a random sample. Perhaps the students who are doing poorly felt they didn’t have time to participate. Or conversely, perhaps they participated to try to win approval from the faculty. If this sample is not representative of the population of interest, any results must be questioned.

We are also concerned that the grades given by different faculty members may be different in meaningful ways. Perhaps one faculty member is an especially easy grader. It’s unclear how this might affect the result.

Finally, a major concern is that it is unclear what it means for the student evaluations to be “as good as” the faculty assessment. This is the primary scientific issue, but has not yet been defined in a meaningful way.

Questions for Client There are several questions we would like to ask the client. First, we’d like more information about how the students in the study were found. Perhaps we can gain qualitative information about how representative the sample may be. Secondly, we’d like to know how the students were assigned to the faculty. Were these assignments random? We’d also like to discuss with the client their idea to treat the faculty response as “truth,” especially with respect to the possibility that there may be easy or hard graders.

Finally, a major discussion with the client must be about what it means for student assessment to be “as good as” faculty assessment.

Idea for Analysis: 2x2 table One idea we had was to simplify the data into two levels; one signifying that the student had obtained mastery (perhaps A–B) and one signifying that they had not (perhaps C–F). This could then be used to make 2x2 tables for each category separately.

The most common test with a 2x2 table is the test for independence, but that’s probably not a reasonable thing to do in this case, as the question is not whether student and faculty scores are independent, but instead to measure how associated they might be and whether they are associated enough.

One idea we had to test this was to instead compute the specificity and sensitivity of using the student grade to predict the faculty grade. This has the added benefit of separating the students with good grades and the students with bad grades; this might be especially useful as the goal is show that students have obtained mastery, so if the self-assessment for

those students who do not have satisfactory grades are not accurate, this might be enough to discredit the idea of using student evaluations instead of faculty evaluations.

Another benefit of this analysis is that it is simple to understand and interpret; depending on the client, that may prove to be an especially good reason to choose it.

Idea for Analysis: Logistic/Multinomial Model We also discussed more sophisticated models, such as a logistic or multinomial model. Models could be built for each category separately, or perhaps all the data could be included in one model. A random effect for faculty member could also be included. This would allow us to model lots of information (if it works) but would be difficult to understand and interpret.

Ideas for Display We also brainstormed ideas for displaying the data; one idea was to use a 5x5 table with counts of how many students had that faculty/grade combination; counts on the diagonal would correspond to equal grades, and counts above and below would correspond to students over or under estimating their abilities.

Another idea was to use a version of a parallel plot with line width corresponding to number of students; such as the following. Lines corresponding to students who gave themselves satisfactory grades but the faculty did not are shown in black.

