2012-2013 Annual Program Assessment Report

Please submit report to your department chair or program coordinator, the Associate Dean of your College and the assessment office by Monday, September 30, 2013. You may submit a separate report for each program which conducted assessment activities.

College: *Science & Mathematics*

Department: *Geological Sciences*

Program: *B.S. (all options)*

Assessment liaison: *Matthew d'Alessio*

1. **Overview of Annual Assessment Project(s).** Provide a brief overview of this year's assessment plan and process.
   - **Completed our "Second Annual Geologic Problem Solving Night", our closed-loop assessment process that tests SLO 3:** demonstrate skills in standard data-gathering and data-analysis methods in both lab and field settings, **with great buy-in from students and faculty.**
   - **Did further analysis, discussion, and improvement of our geoscience concept inventory to test SLO 1:** demonstrate conceptual understanding of different earth materials and the processes that shape them throughout their history**.**
   - **Finalized the overall structure of our curriculum redesign for the B.S. program. Individual courses are laid out in relation to one another and in relation to the skills we define in our program SLO's.**
   - **Created a new rubric and assessment process for project proposals, an authentic assessment of SLO 4:** present polished summaries, both written and oral, of their geological discoveries. (not discussed in this year's assessment report. Stay tuned for next year)
   - **Hosted an assessment-focused speaker for our department colloquium (Dr. Marcy Osgood), including two days of faculty discussions with the speaker who is an expert in problem-solving skill development and assessment.**

2. **Assessment Buy-In.** Describe how your chair and faculty were involved in assessment related activities. Did department meetings include discussion of student learning assessment in a manner that included the department faculty as a whole?

Department chair: Our chair ensures that timetables are met and adequate time is devoted to discussion at faculty meetings. She allocates adequate release time to the assessment liaison and ensures that the proper tone is set regarding the role of assessment in the department.

Faculty contributions to assessment design: Assessment instruments and rubrics are typically created by the liaison and then sent to the faculty for comment over email. However, the development of the assessment rubric and process for project proposals was driven by the course instructor of research design. The liason strongly supported this effort, and several other faculty also contributed. In the last year, half our faculty (6 out of 12) have replied with substantive comments about one or more assessment design issues, though any individual query I send usually receives feedback from about a third of the department (4 out of 12 faculty). These contributions are not trivial and often consist of actually completing an entire assessment as a 'student' so that I can get 'expert-like' answers as models and then offering review comments in addition. While not everyone participates, there is enough critical mass that we get things done.

Faculty contributions to assessment implementation: Faculty readily embed assessment items in their courses. In fact, they often check in with the assessment liaison and ask, "shouldn't I be using an assessment item in my course soon?"

Faculty contributions to assessment scoring: Ten of twelve faculty participated in scoring proposals by our graduate students and senior thesis candidates, with each proposal read by three faculty. These were multi-page proposals and faculty scored them using rating scales for 20 different criteria along with providing specific feedback on those criteria.

Faculty discussions of assessment data: Last year, we began our annual “Geologic Problem Solving Night” where our undergraduates walk through the scientific process through a progressively revealed problem. We devote a substantial fraction of one faculty meeting to reviewing the responses of our students, scoring them, and discussing implications. This year, we had 9 of 12 faculty participate in this exercise (two faculty were unavailable on medical and parental leave). Our faculty are particularly interested in addressing our weaknesses as we rewrite the undergraduate curriculum in our program (we expect to submit to the College program modification paperwork in Spring 2014).

SLO's Asssessed

3. **Student Learning Outcome Assessment Project.** Answer items a-f for each SLO assessed this year. If you assessed an additional SLO, copy and paste items a-f below, BEFORE you answer them here, to provide additional reporting space.

3a. Which Student Learning Outcome was measured this year?

1) demonstrate conceptual understanding of different earth materials and the processes that shape them throughout their history;

3b. Does this learning outcome align with one or more of the university's Big 5 Competencies? (Delete any which do not apply)

NO. This is discipline-specific content knowledge.

3c. Does this learning outcome align with University's commitment to supporting diversity through the cultivation and exchange of a wide variety of ideas and points of view? In what ways did the assessed SLO incorporate diverse perspectives related to race, ethnic/cultural identity/cultural orientations, religion, sexual orientation, gender/gender identity, disability, socio-economic status, veteran status, national origin, age, language, and employment rank?

NO.

3d. What direct and/or indirect instrument(s) were used to measure this SLO?

Last year, our department developed our own version of the Geoscience Concept Inventory, a national test designed to assess conceptual understanding of some big-picture ideas in Earth Science. The assessment consists of 24 multiple choice items probing possible misconceptions students have about the most important principles of geoscience. Questions on the GCI are well validated for introductory geoscience students, but our application is unique in the context of geoscience majors. As we discuss below, the additional knowledge possessed by our majors makes some test items invalid because the students understand more aspects and complexities of how distractors may be partially correct, which can lead to multiple keys on a multiple choice item.

3e. Describe the assessment design methodology: For example, was this SLO assessed longitudinally (same students at different points) or was a cross-sectional comparison used (Comparing freshmen with seniors)? If so, describe the assessment points used.

We performed a cross-sectional comparison between students in our gateway course for majors (GEOL 207) and an upper-division required course that is typically the last required course before students take electives (GEOL 310). We do not have a true capstone course in our major, but we refer to GEOL 310 as our capstone for the purposes of this report. We also have pilot data from a GE-level course from two years ago (GEOL 106LRS, n=105). This year's report extends last year's report, which assessed the same SLO. We administered the instrument again to increase our sample size and to discuss the results in more detail.

3f. Assessment Results & Analysis of this SLO: Provide a summary of how the results were analyzed and highlight findings from the collected evidence.

GE-Level growth. Our data show that GE students do leave our classes with improved understanding of geoscience concepts, but we have a long way to go. Our results are comparable to scores on this instrument nation-wide (Libarkin & Anderson, 2005).

GE-to-majors transition. We recruit a number of majors from our GE classes, and all majors are required to take at least two GE classes. The significant difference in means is due to selection effects. We are much more likely to recruit students into our program that do well in the GE class.

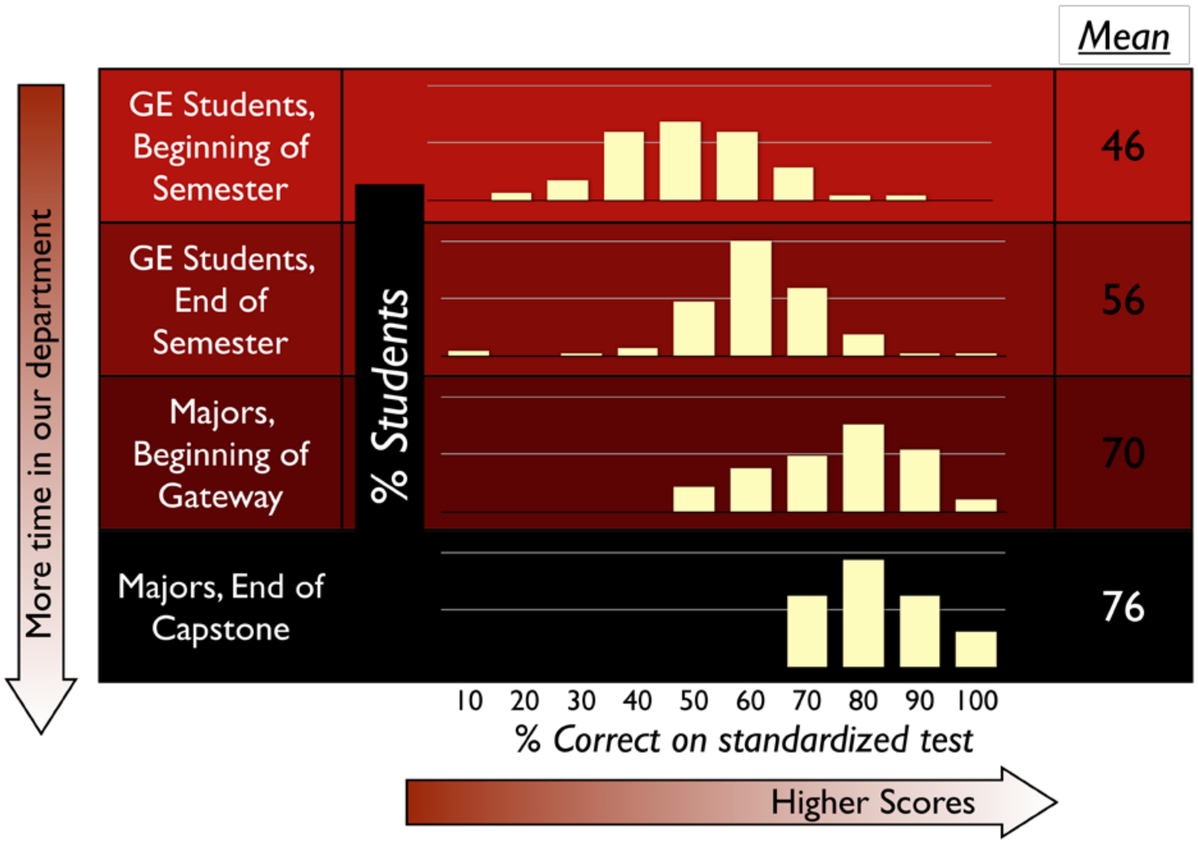
Gateway repeatability. Results from last year and this year for GEOL 207 are comparable and statistically identical (means of 69.3±14.4 v. 71.4±12.8). As our major is growing, we now have gateway scores for 46 students.

Gateway to Capstone improvement.
Qualitatively, there are large differences in the data. A third of the students in GEOL 207 have scores of 65% or lower. By contrast, the lowest score in GEOL 310 was a 68% – not a single student scored 65% or below. We interpret this finding to mean that we are doing a good job helping our low-performing students increase up to the level of our higher achievers. Mean scores, however, do not shift much (with the mean, median, and mode scores in the 70-80% bin for both groups). This indicates that we do not address enough of the deficiencies in students that start with more background understanding. The item-by-item analysis we completed in last year's assessment report shed light onto subjects included in this instrument that our program does not adequately explore (such as global climate), so we would not expect growth in any of the students on those items (action items below to solve this). There are also questions on the nationally validated instrument (~8% of the total) that our faculty feel are flawed items and we therefore would not expect growth by any students on those items (action items below to solve this).

Because of the low sample size (only 8 students in GEOL 310 – an artifact of a smaller program in the past), it is unwise draw sweeping conclusions from the qualitative data. While the qualitative difference was dramatic, the statistical difference is smaller.

Students in GEOL 310 do better than GEOL 207 (means of 76.2±8.5 v. 70.2±13.7), and the difference is almost statistically significant (different with 94.9% probability using a one-sided t-test. A single additional data point from GEOL 310 would give us enough resolving power to cross the 95% threshold). However, using statistics to argue that there is no difference would be misleading because the skewed tail of low scores in GEOL 207 violates the statistical assumption of a normal distribution.

We will repeat the measurements again in 2013-14. Next year's data will give us a sample size big enough to make conclusions from the cross-sections and a chance to longitudinally examine the growth of our first cohort from GEOL 207.



3g. Use of Assessment Results of this SLO: Describe how assessment results were used to improve student learning. Were assessment results from previous years or from this year used to make program changes in this reporting year? (Possible changes include: changes to course content/topics covered, changes to course sequence, additions/deletions of courses in program, changes in pedagogy, changes to student advisement, changes to student support services, revisions to program SLOs, new or revised assessment instruments, other academic programmatic changes, and changes to the assessment plan.)

New course that addresses gaps. Last year, we found that students did particularly poorly at items related to Earth's climate. We recognized this is a weakness in our curriculum. To address this missing piece during our curriculum redesign, we are creating a new 200-level course called Earth System Science. While it will draw largely from content in our existing entry level classes related to solid-Earth processes, it provides the framework for addressing interactions between Earth's climate and other geologic systems. As planned, 2-3 weeks of that course will introduce climate, focusing on climate science and global climate change.

Instrument improvements. Last year, the faculty identified two problematic items in the nationally validated multiple choice test. We decided to leave them as-is for another year, but students continue to universally select a distractor because it is partially correct. While the test is specifically designed to capture misconceptions, a number of our faculty chose the distractor as well. We reworded the items so that they still identify the misconception without risking multiple keys. Next year, we expect to see dramatic changes in the rate of correct answers for these two items (which, out of 24 items, represents a substantial change in average score).

---

3. Student Learning Outcome Assessment Project. Answer items a-f for each SLO assessed this year. If you assessed an additional SLO, copy and paste items a-f below, BEFORE you answer them here, to provide additional reporting space.

3a. Which Student Learning Outcome was measured this year?

3) **demonstrate skills in** standard data-gathering and **data-analysis method**s in both lab and field settings

3b. Does this learning outcome align with one or more of the university's Big 5 Competencies? (Delete any which do not apply)

- Critical Thinking

- Quantitative Literacy

3c. Does this learning outcome align with University's commitment to supporting diversity through the cultivation and exchange of a wide variety of ideas and points of view? In what ways did the assessed SLO incorporate diverse perspectives related to race, ethnic/cultural identity/cultural orientations, religion, sexual orientation, gender/gender identity, disability, socio-economic status, veteran status, national origin, age, language, and employment rank?

Not really. The geologic problems we use are socially relevant. This year's topic was global climate change and utilized data from cherry blossom bloom dates in Japan. It briefly introduced students to the Japanese cultural phenomenon of *hanami* through description and pictures (and a data set).

3d. What direct and/or indirect instrument(s) were used to measure this SLO?

We developed this methodology in last year's assessment activities. We utilized a "progressively revealed problem-solving assessment." It is based on some of the case study assessments done by the Biology department at University of New Mexico (Anderson, 2001). In order to effectively implement this assessment strategy, our department colloquium hosted Dr. Marcy Osgood, one of the authors of that paper, on November 1, 2012. Faculty met with Dr. Osgood to specifically discuss this assessment strategy and strategies her group uses for promoting problem-solving skill development. We spent two days with groups of faculty meeting her for discussions over a brow bag lunch, dinner, and breakfast (as well as in offices and during her formal research talk). In total, eight of our twelve faculty met with Dr. Osgood during her visit.

The analysis tests 5 independent skills in scientific problem solving:

1. Hypothesis generation
2. Planning investigations
3. Analyzing data (*quantitative literacy*)
4. Evaluate Conclusions (*critical thinking*)
5. Reflect on Conclusions and Formulate Next Steps

We present students with a single scenario ("The Case of Blooming Sakura" for this year's prompt). They respond to the 5 key skills listed above in a series of 5 prompts. With each new prompt, new information is revealed. While all build on a single scenario, the information required for each skill is provided within each skill's prompt. In other words, students are not penalized for "not-knowing how to answer a previous question" because a valid correct answer to the previous question is revealed each time the student moves forward. The situation is novel enough that no students have prior knowledge tackling this specific problem. Though they certainly do draw on some background knowledge in geology, we are hoping to test these independent scientific process skills. We implemented the exercise electronically in Moodle.

Last year, we found that our students were excellent at hypothesis generation (including simultaneous discussion of multiple hypotheses) and lacked skills in data analysis, interpretation, and evaluating conclusions. This year, we saved time by skipping the hypothesis stage and adding a preliminary data analysis stage to probe explanations in this area more deeply without adding extra length to the problem session.

Here is how we implemented this system:

1. After an introduction to *hanami*, students view a plot with cherry blossom bloom dates over the last 1000 years. Students are asked to simply narrate what they see in the simple data set. This gives them familiarity with the problem and primes them for thinking about this natural system that they have probably never considered as geologists.
2. Upon submitting that response, they are asked to pursue one specific hypothesis chosen by the exam writers (it may not necessarily have occurred to the student as they examined the data). To test their ability to plan the investigation, they enter in a list of different types of relevant data that would help them test the given hypothesis.
3. After hitting submit, they are presented with actual data chosen by the exam writers (again, it may not coincide with any of the data they requested in the previous step). Students are asked to narrate an analysis of the graphs, looking for trends and relationships. They must also decide whether or not their original hypothesis is supported by the data set.
4. After hitting submit, they are asked to defend a specific conclusion chosen by the exam writers. They enter in a response to a fictitious anonymous peer reviewer defending their position and/or acknowledging the viewpoints in the review comments.
5. After hitting submit, they are asked to reflect on the entirety of the scenario thus far. What new questions do they have that they would like to pursue based on the data and analysis they've seen?

Responses to each prompt are relatively short paragraphs or bullet points.

3e. Describe the assessment design methodology: For example, was this SLO assessed longitudinally (same students at different points) or was a cross-sectional comparison used (Comparing freshmen with seniors)? If so, describe the assessment points used.

Lacking a formal capstone course in our major, we built on the success of our assessment process from last year. We sponsored the "Second Annual Geologic Problem Solving Night." We invited all our undergraduate majors to attend an evening of free pizza and ice cream. We asked faculty NOT to offer extra credit for participation. Instead, we asked faculty in our core undergraduate courses to provide one question from their upcoming final exam. We would reveal the question to students and discuss the answer. This was a STRONG motivator to students, and we even had graduate students enrolled in these core classes ask if they could attend, too.

Students sat at laptops and worked individually on the progressively-revealed problem solving exercise on Moodle. Considering this was an evening activity the week before final exams, we were thrilled by the results! Twenty-one students volunteered to show up and take the assessment. While working through the problems, students were absolutely silent and focused. Several students had just completed culminating presentations of their senior theses earlier in the day but still chose to participate. The quality and seriousness of their responses are a testament to this focus. We feel that this experience is a strong counter-example to the idea that students need to be graded on assessment items in order for them to take them seriously.

From the perspective of rigor, our sample is highly biased. The participants were self-selected. We can state that they come from a wide range of backgrounds and class-levels (approximately equally from sophomore, junior, and graduating senior level in our program).

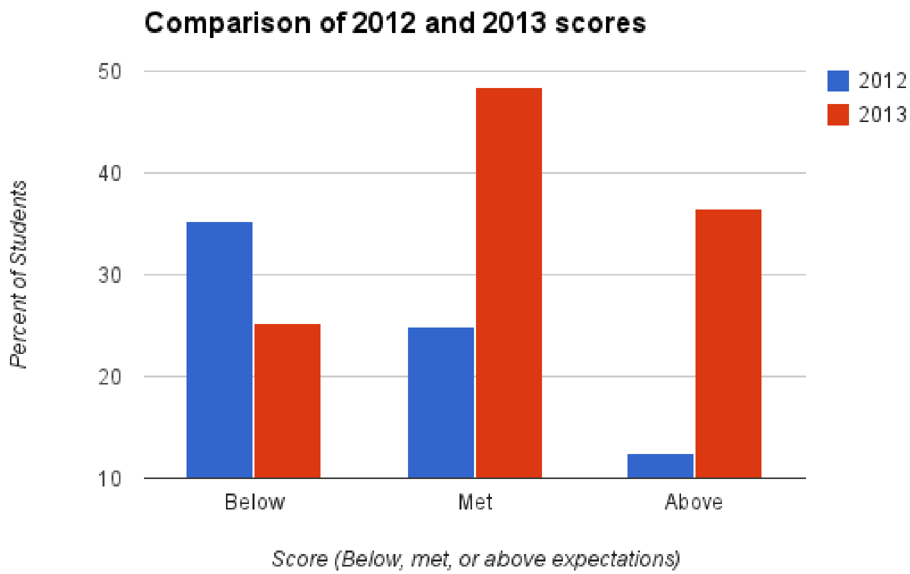
3f. Assessment Results & Analysis of this SLO: Provide a summary of how the results were analyzed and highlight findings from the collected evidence.

Our major order of business at the first faculty meeting of the semester was to analyze and discuss the Annual Problem Solving results. Faculty members worked in pairs (or groups of 3) to consider each of the independent sub-skills. I presented each pair with the definition of each skill and a provided a copy of the student prompt. Before they were allowed to view student responses, I asked them to formulate their own response and think about what they expected of undergraduates in our department. After discussing this holistic 'rubric' with another faculty member, they were allowed to evaluate student responses, adjusting as necessary after seeing a sample of student responses. Each response was scored using a three point scale:

1. Below expectations for an undergraduate about to receive a B.S. in our department
2. Met expectations
3. Above expectations

Since the responses are only a few sentences each, faculty can read all the responses, score a set individually, discuss the set with their partner to produce a consensus score, and continue on to score the next group of responses. They quickly enter their scores in a collaborative spreadsheet so that we can see histograms of performance of each cohort of our students broken out by scientific skill. In other words, the entire faculty participates in the process of developing the specific holistic rubric for their item, norming the rubric, scoring all responses in that category, and discussing the corpus of results with fellow faculty. It’s certainly not a perfect process, but it is an extremely efficient way for faculty to be involved in the process and generates a lively discussion about the implications of our findings.

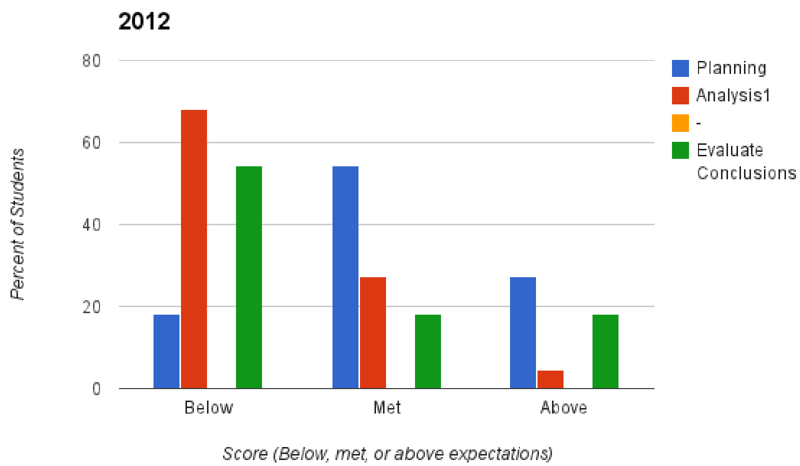
The data faculty saw in the meeting are below:

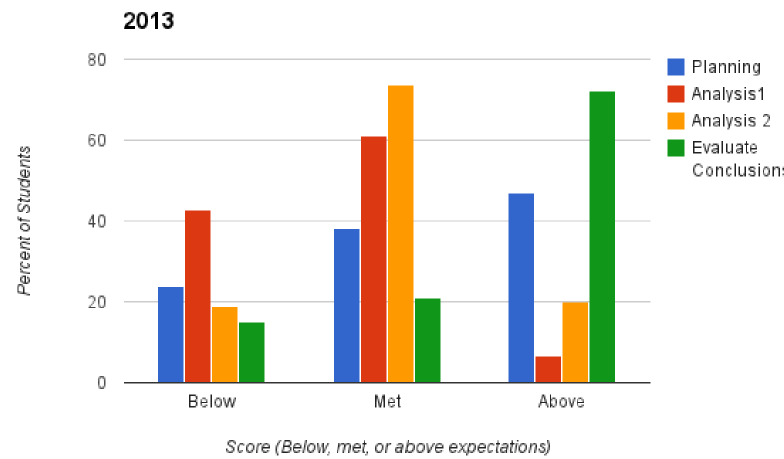


There are obvious differences between 2012 and 2013 that likely relate to the differences between the two challenge problems used each year (while similar in set up, they were likely not equivalent in difficulty) or inconsistencies in scoring on the part of faculty. We do *not* interpret these differences to any change in performance from one year to the next reflecting substantial improvement to our program. Seven out of 21 students completed the assessment both years, allowing for a longitudinal comparison in principle. However, the dramatic differences in scores between the two years suggest that this would be an unwise analysis. (All 7 students showed substantial 'improvement' in scores from 2012 to 2013, but the improvement was no different than the overall mean difference between the two years).

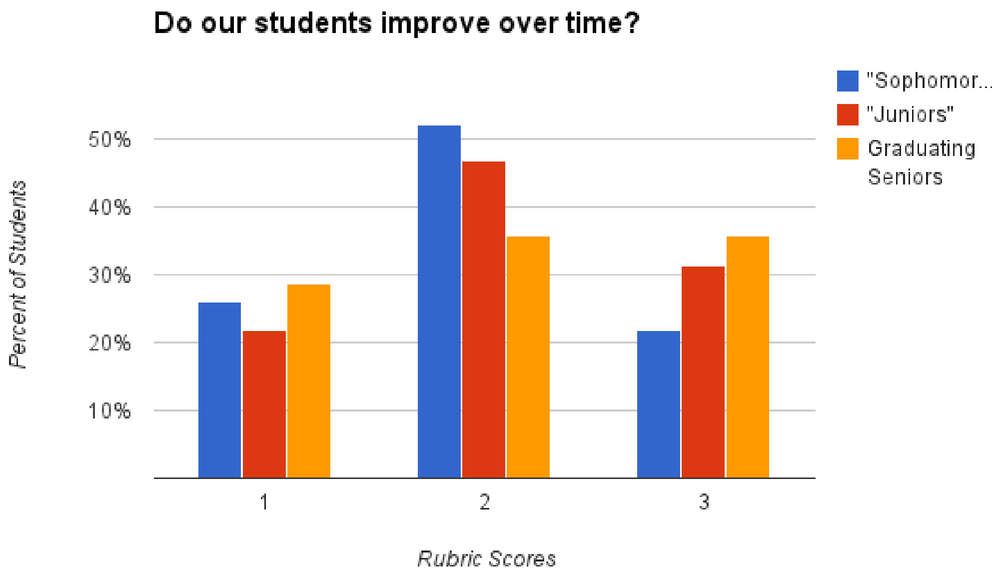
This year, we assessed data analysis twice for each student and can evaluate the reliability of our instrument by comparing the two scores student-by-student. In aggregate, the mean scores were 1.7 and 2.0 for these two sections. Since each section was scored by different faculty but considered portions of the exact same data set, we attribute the different mean scores largely to inconsistency in scoring. For individual students, scores were identical on the two items only 30% of the time, differed by one rating level 50% of the time, and differed by a full two levels 20% of the time. While both scoring groups agree that there were less than 20% of the responses that were above expectations, the aggregate scores differ in the fraction of students below expectations (20% versus 40%; see red and yellow bars below).

As far as accuracy of the scoring, faculty were concerned that the high level of scores above expectations in the category "evaluating conclusions" may be attributed to low expectations and easy scoring on the part of the faculty group in charge of that item. Both faculty scorers in that category left the meeting a bit early and were not present to discuss the results. In 2012, 55% of our students were below expectations in this category and in 2013 72% are above expectations. Since no dramatic changes were made in our curriculum, these ratings may be inaccurate.





Despite these issues with reliability and accuracy of the instrument, one more piece of data helps us gain a little trust in the overall picture the data show. Scores that exceeded our expectations were lowest in our sophomore group and highest in our graduating senior group, which we see as a relatively robust measure of growth over time. There is some concern that the largest fraction of students below expectations also come from our graduating seniors, but faculty were inclined to attribute this result to “senioritis” as students did not care as much about producing detailed explanations. I disagree – three of the eight seniors that participated had literally just completed their senior thesis presentations moments before this assessment exercise. The relief among these students of having completed such a major milestone would make these students most prone to senioritis, but only one single item out of the 12 we scored from this group was below our expectation (8% versus 35% for the non-thesis seniors doing the assessment). The students that complete senior thesis are among our best students, and their assessment scores in this exercise were indeed the highest.



3g. Use of Assessment Results of this SLO: Describe how assessment results were used to improve student learning. Were assessment results from previous years or from this year used to make program changes in this reporting year? (Possible changes include: changes to course content/topics covered, changes to course sequence, additions/deletions of courses in program, changes in pedagogy, changes to student advisement, changes to student support services, revisions to program SLOs, new or revised assessment instruments, other academic programmatic changes, and changes to the assessment plan.)

Improving data analysis skills. Our results show a consistent weakness in data analysis and interpretation of complex data sets in both 2012 and 2013. I do not feel that the problem is lack of exposure to data or data analysis and interpretation tasks. Interpreting data like an expert, like all expert skills, takes time and experience to develop (e.g., d'Alessio & Lundquist, 2013). While we provide opportunities for students to attempt to interpret data, I think we can definitely improve on our systems for giving feedback to students about their analysis. Our faculty claim to give detailed feedback on lab reports, which they can do because of our relatively small class sizes. However, is this feedback timely enough? Are there structures in place for students to actually respond to the feedback and act on it so that they internalize it? For example, a lab report returned a week later with comments and a summative grade does not necessarily promote the growth we want. Some faculty mention that they allow students to do exam corrections for extra credit, but it's possible to build in formal structures where there is the expectation that all products (especially data analysisrich ones) will be iteratively improved in response to feedback. This process might involve more small assignments with feedback and fewer 'authentic' assignments that require in-depth data collection and lengthy analysis. The faculty need to consider some options and determine how they can be implemented without huge growth in the grading/feedback time obligations. As we implement our newly redesigned courses, we need to devote attention to these structures. At present, no action was taken by the faculty beyond the thought-provoking discussion. This discussion is a continuation of the discussions we had with Dr. Osgood in November. During the coming academic year, we will target this discussion towards specific activities and strategies that faculty plan to use in the new course redesign.

Changes to assessment process. While I see enormous value to having the faculty participate in this mass-scoring faculty meeting where pairs of faculty read responses and rate them, the faculty mentioned that having a single scorer read all responses and score them would probably be the best way to ensure reliable scoring and perhaps better quality data. Next year, a small group of faculty volunteers will complete this task and report the consistent scoring results to the faculty. This will not replace our faculty meeting, but will be in addition to it (I feel it is valuable for faculty buy-in and to give them a real hands-on sense of the state of the students).

---

4. Assessment of Previous Changes: Present documentation that demonstrates how the previous changes in the program resulted in improved student learning.

No documented improvement to-date.

5. Changes to SLOs? Please attach an updated course alignment matrix if any changes were made. (Refer to the Curriculum Alignment Matrix Template, http://www.csun.edu/assessment/forms_guides.html.)

No changes made to SLO's.

6. Assessment Plan: Evaluate the effectiveness of your 5 year assessment plan. How well did it inform and guide your assessment work this academic year? What process is used to develop/update the 5 year assessment plan? Please attach an updated 5 year assessment plan for 2013-2018. (Refer to Five Year Planning Template, plan B or C, http://www.csun.edu/assessment/forms_guides.html.)

We have been on-task with the 5 year plan, largely because our assessment discussions continue to be driven by our large curriculum redesign project (which has a five year funding horizon). Our biggest challenge for the next 3 years is implementing improvements. The faculty have been very willing to meet and discuss the overall design of the program to ensure it is integrated, but may be less willing to sit down and discuss effective strategies for implementing that plan. There are disagreements about teaching styles and approaches and we all consider ourselves experts on the delivery of our specific content based on years of experience.

7. Has someone in your program completed, submitted or published a manuscript which uses or describes assessment activities in your program? Please provide citation or discuss.

My recent publication listed below is not directly related to our current assessment data in Geology, but it informs our thinking about data analysis and how to train students to analyze data like experts:

d'Alessio, M., & Lundquist, L. (2013). Computer Supported Collaborative Rocketry: Teaching students to distinguish good and bad data like an expert physicist. The Physics Teacher, 51 (7), 424-427, doi: 10.1119/1.4820858 <Link to journal>

8. **Other information, assessment or reflective activities or processes not captured above.**

References cited

Libarkin, J.C., and Anderson, S.W., 2005, Assessment of learning in entry-level geoscience courses: Results from the Geoscience Concept Inventory: Journal of Geoscience Education, v. 53, p. 394-401.

Anderson, W.L., Sensibaugh, C.A., Osgood, M.P., & Mitchell, S.M. (2011). What Really Matters: Assessing Individual Problem-Solving Performance in the Context of Biological Sciences. International Journal for the Scholarship of Teaching and Learning, 5(1). Retrieved from http://academics.georgiasouthern.edu/ijsotl/v5n1.html