Selecting which outcome and approach

In September 2016 the Core Advisory Council (CAC) selected the undergraduate requirement of Science plus Laboratory as the focus area to assess student learning this year.

The Council noted a few driving factors for the selection:

= almost all students fill that requirement by taking a course outside their major;

= current policy requires transfer students to bring only a science lecture to satisfy the core, with no lab required;

= non-science-major lab sections do not require math prerequisites, hence limiting the kind of quantitative work that can be done in lab.

Three of the ten undergraduate core outcome statements are relevant to the Science plus Lab course requirement:

> 1. Critical and Creative Thinking: to analyze information logically and to utilize and transform knowledge in fair-minded, purposeful, and imaginative ways.
>
> 7. Quantitative/Scientific Analysis: to use mathematical reasoning and the scientific method to address issues in an increasingly complex, technological world.
>
> 10. Integration/Spirituality: to understand the importance of developing as a whole person who is spiritually mature and dedicated to being a productive and responsible citizen.

The third core outcome statement is relevant because many core science courses include an outcome that addresses student awareness of human impact on natural systems and the global environment.

Science plus Lab Assessment Working Group

In consultation with the SMSE Dean, Dr. Garcia, and CAC members, a faculty working group was formed to design the assessment of student learning in Science plus Lab courses.

> Dr. Sara Tallarovic, CAC, Biodiversity course coordinator
> Dr. Ric Peigler, Diversity of Life course coordinator
> Dr. Mike Maguigan, Intro to Environmental Science course coordinator
> Dr. Zhanbo Yang, chair for Physics and Environmental Science
> David Stein, CAC, EAP assessment director
> Dr. Glenn James, CAC co-chair

Starting in December, the working group met once a month to design and implement the assessment project for the year.

Resources and Rubrics

The working group examined the list of all the courses currently approved to satisfy the Science plus Lab requirement, and took into account the typical enrollments in those courses (see Atch 1 for numbers). Courses were considered for assessment by applying the following criteria and goals for the project:

= pick courses with plenty of students and plenty of non-science majors
= be able to compare findings from multiple courses
= aim to compare main campus, adult learning, and CIW Mexico city students

The group looked at learning outcomes on current syllabi for those courses, and narrowed down two outcomes that were common on nearly all syllabi:

(a) Demonstrate effective, structured problem solving.
(b) Evaluate the role/impact of humans in the relevant course topic.

The overall strategy was to follow recommendations from Dr. Catherine Wehlburg (TCU) – and apply UIW experience from previous years of core assessment – and design a general core-learning rubric that can be applied to existing coursework in multiple courses. This approach minimizes the intrusion on faculty and courses, and assures that assessments are done with direct measures of student work.

Draft input for rubrics came from multiple VALUE Rubrics (Lifelong Learning, Critical Thinking, Quantitative Literacy, Inquiry and Analysis) and comparable published samples from other institutions. A long-form rubric was composed that can serve as a starter rubric for almost any quantitative course (Atch 2). A short-form with only four outcome rows was narrowed down to outcomes most appropriate for introductory-level courses, across disciplines (Atch 3):

Application of knowledge to hypothesis
Data interpretation
Problem identification
Role of Humans

Courses and assignments collected

The working group invited faculty from several courses and disciplines to contribute graded collections of papers, exams, and projects that the instructors felt might demonstrate student learning in at least one of the four rubric outcomes. CIW Mexico City faculty confirmed that they ran a section of Diversity of Life in Spring 2017, which worked out perfectly, since Diversity is the largest enrolled science course on main campus, and was the obvious primary source of student samples for the assessment.

In the end, multiple faculty from three courses provided student samples from an existing assignment where learning could be assessed in one or both of the "common" outcomes (a) and (b) listed above. Most instructors provided multiple-choice Scantron tests; some provided short-answer, and one provided slides from student presentations.

Ultimately over 270 samples were collected from the following courses:

BIOL 1401 Diversity of Life	111 samples (23 from adult ed; 6 from Mexico)
BIOL 1403 Biodiversity	125 samples
ENSC 1410 Intro to Environmental Sci	43 samples

Scoring Day

The scoring team included working group members, David Stein (facilitating), Dr. Ric Peigler, Dr. Zhanbo Yang, and Dr. Mike Maguigan, with the additional help of senior adjunct Biology faculty member, Dr. Tom Stedman. Drs. Peigler and Stedman brought the essential insight of instruction and leadership in Diversity of Life; Drs Maguigan and Yang are the faculty leaders for Environmental Science.

The team met and scored on a single day, during summer break.

Much time and effort was devoted in the weeks prior to the meeting to develop the scoring rubric and collect student samples. Additional effort was spent in the days leading up to the meeting to redact student and faculty names, as well as other unique identifiers, from the samples. A short overview of this process, and the general plan for the day, was presented by the meeting facilitator and agreed to by the group.

Two sets of student samples were then distributed for discussion. Based on our prior years' scoring experience with core learning student samples, we emphasized that the assessment effort was about *student learning* more so than instruction. We were there to tease out strengths and weaknesses from the student samples to evaluate the extent to which our students were acquiring knowledge and developing associated skills. In an effort to help inform future institution-wide assessment efforts, the group was also encouraged to reflect on the assessment process itself.

The initial guidance to the group was to team up with a colleague so that each student sample could ultimately be read and scored by two people. Then they were asked to carefully review the student samples in an effort to identify particular prompts (from the exam instruments) that were most relevant to the four outcomes on the scoring rubric. Once this had been accomplished, each pair was asked to develop a method that would allow them to determine the proficiency on the scoring rubric – Proficient (3), Basic (2), Unsatisfactory (1), or Not Applicable.

The two working group pairs collaborated on the final steps of this process. Since several of the student sample sets involved long multiple-choice final exams, the choice of which questions to use to assess student learning proved to be the primary challenge. The readers ultimately decided to pick their "best three" questions, when they encountered exams where ten questions or more might specifically address one or more of the four rubric outcomes. This approach proved very workable. It also made inter-rater reliability a non-issue: Once the prompts to be used in the assessment were chosen, there was no room for disagreement as to the score required on the rubric. Students who answered all three of the prompts correctly were assessed as being Proficient (3) in that area. Those who answered two of three correctly were scored as having a Basic (2) level of understanding. All others were deemed to be Unsatisfactory (1). In several instances, the readers found that the prompts given to students did not relate to one or more of the four component parts of the scoring rubric and were, therefore, scored as "Not Applicable."

Scores from the rubric were recorded for each student sample and then verified by both scorers. Maps were created for each set of student samples showing which questions on the exams were used to assess each of the four component parts of the scoring rubric. Where necessary, other notes were also recorded to identify potential issues of concern. In one case, for example, a note was made regarding the exclusion of a small group of students from one of the samples. Their scores on other assignments during the term were so high that they were not required to take the test that was being used for this assessment effort.

A week after the scoring meeting, Stein and James met to review the scoring session. They also collated and scored two additional samples of student work from two sections of BIOL1401 – Diversity of Life taught in the recently completed EAP summer term. One of these samples was a collection of PowerPoint presentations. The other came from two short-answer questions from a final exam.

For the Mexico City section of Diversity of Life, the class had only had six students this term. The instructor coordinated closely with the UIW working group, and that instructor used the identical rubric to assess the student learning in the specific rubric areas. The assignment assessed was a group project, and while there were *graded* elements that were different for the six students, the *rubric* areas aligned with scores that went to the entire group, so all six Mexico City results were identical for all six students.

Data Analysis

It is important to note that the scoring group was not asked to perform the kind of in-depth analytical research that would otherwise be required to publish a related article. Rather, they were asked to review the data in a way that would allow them to identify related strengths and weaknesses that would, in turn, inform continuous improvement efforts.

It must also be noted that individual scores used in this exercise were not the same for different sections of student samples. For instance, though a score of "3" most often meant that a student answered three comparable multiple-choice questions correctly, it occasionally meant that they answered only *one* question correctly. We used the scoring simply to observe strengths and weaknesses in student learning, not to gather precise statistics for these samples.

A total of 279 UIW student samples, taken from 14 different sections of classes offered in Spring and Summer of 2017, were collected and analyzed for this assessment effort. Approximately 90% of these were from classes offered on the main campus while the remaining 10% were from EAP and the campus in Mexico.

The student samples generated a total of 636 individual scores in the four areas assessed by the rubric. Most of the assignments used to collect student responses allowed for scoring only in one or two of the four areas. The scores ranged from 3 (Proficient) to 1 (Unsatisfactory).

The following table provides summary data collected from the assessment.

	Apply Knowledge	Data Interpretation	Problem ID	Human Impact
Overall Averages	2.42	2.23	2.40	1.97
Overall Variance	0.55	0.75	0.51	0.64
Sample Size	153	168	119	196
Scores of 3 - Proficient	88	86	64	59
Scores of 2 - Basic	42	34	39	74
Scores of 1 - Unsat	23	48	16	62
Number of Total Scores	153	168	119	196
Percent 2 or Above	85.0%	71.4%	86.6%	68.4%
Percent Less than 2 (Unsat)	15.0%	28.6%	13.4%	31.6%
Diversity 1401 (non-Majors)	1.57	2.18	2.11	2.01
Biodiversity 1403 (Majors)	2.60	2.32	2.51	1.90
Environmental Science (non-Majors)	N/A	N/A	2.18	1.81

Findings

The overall results of the assessment effort were promising, suggesting that most UIW students are acquiring at least a basic level of knowledge and skill in each of the four areas assessed by the rubric. Of these, Applying Knowledge (2.42) and Problem Identification (2.40) were strongest. Human Impact (1.97) was weakest. Approximately 30% of the scores generated for Human Impact and Data Interpretation were Unsatisfactory.

When the data set was segmented by majors, the average scores for science majors were generally higher than those for non-science majors in three of the four rubric areas. The one exception to this finding was the fourth area, Human Impact. A closer look at the prompts used to evaluate this area suggested that the difference was due in large part to a higher level of expectations for those science majors. In particular, the test questions in that area, for the majors' students, called for more analysis and synthesis (higher Bloom's taxonomy) than other recognition-level questions for non-majors in that outcome.

Additional findings:

1. Though sample sizes were relatively small, scores from CIW Mexico City were generally stronger than average while those from EAP were weaker than average.

2. The use of student samples rooted in multiple choice instruments or PowerPoint presentations did not pose assessment challenges. Inter-rater reliability was guaranteed for the samples that relied on multiple-choice questions. Where samples were composed of PowerPoint presentations or free response final exam questions, the two scorers generated evaluations with correlations (inter-rater reliability) over 90%. Nearly 76% of these pairs of scores were identical and another 21% were different by less than a point.

3. The only meaningful evaluative challenge occurred when scoring a single set of responses generated from the last question among 10 open ended questions on a final exam. The prompt itself was reasonable, however its positioning at the end of a lengthy exam generated responses that were generally too brief to be an adequate gauge of student learning. The scorers therefore decided it best to pair responses from this question with another relevant question (#8) on the same final exam.

4. The time invested early on developing the scoring rubric yielded dividends. Its relative simplicity, measuring just four areas and using a three point Likert-scale, afforded the flexibility needed to assess samples from a wide range of UIW classes and assignments.

Final Rubric and Recommendations

General Findings

A. The working group was quite satisfied with the four-outcome rubric, and the members recommend sharing it for wider use, and adaptation to individual courses as needed.

B. Among the four outcomes that were assessed, the baseline student proficiencies in Data Interpretation and in Human Impact only averaged 70%. The faculty reps on the working group will meet with their respective departments and determine if that is a satisfactory baseline, or if they expect a higher average standard as a target.

C. [updated in v11 of this report] The faculty working group found that the student samples from A&P I did not include learning outcomes about Scientific Method or about the impact of humans on the environment. Subsequent review from department faculty confirmed that additional outcomes for Critical and Creative thinking and for Quantitative / Scientific analysis are assessed in lab assignments, which are taught in sections separate from the lectures.

Improvements in Courses

D. Faculty in Diversity of Life, and in Biodiversity, noted that all sections already incorporate a petri dish experiment where students swab multiple public surfaces, hypothesize where the most germs will come from, and observe and report the results. One recommended improvement, from the assessment, is to re-emphasize for faculty the core outcomes being taught in that exercise, and assure they take advantage of the lab to assess the outcomes of:

- Problem identification
- Application of knowledge to hypothesis
- Data interpretation

E. Diversity of Life faculty also will integrate parallel initiatives (from QEP participation) to add writing elements to the course, and design the writing activities to address the core outcome of Human Impact. This step will assure that all sections of Diversity address that outcome with a comparable approach.

F. Assessment for Intro to Environmental Science showed that all course sections were adequately presenting the four outcomes, and student learning was adequate. The improvement proposed was to add short writing elements – such as “one-minute essays” – to reinforce learning in the outcome of Problem Identification which had slight weaker scores.

G. At CIW Mexico City, student learning in the steps of Scientific Method were strong. At the same time, the faculty there recognized that Human Impact was not currently addressed. The topic will be added to the standard curriculum.

Improvement in Assessment

H. The working group agreed that there would be great value to raise overall faculty awareness of the ten core learning outcomes, and especially the four working outcomes that were assessed in this report. They asked Dr. James to write a summary memo that we can share, especially with off-campus and part-time faculty members.

NEXT STEPS

This was a first-assessment of lower-division student learning for these outcomes. The results provide excellent baseline data and indicators for freshman and sophomore learning.

Two next steps are essential. The first is to assess student learning in later courses, and look for evidence of improved learning in more advanced students. Second, learning needs to be assessed for students who transfer to UIW with a science lecture but no lab, compared to students who take the required Science Plus Lab at UIW.

Dr. James will recommend to Core Advisory Council to do those steps for the coming year's assessment cycle.