

Selecting which outcome and approach

In 2012-13, the Core Advisory Council (CAC) extended the student writing assessment begun by the department of English, and concluded a wide-reaching assessment of graduate student writing. That exercise was the first step in the more deliberate cycle recommended by Dr. Barbara Walvoord at her summer UIW conference (May 2013), where she advised CAC to work more to prioritize its work, and take two years if needed to assess each priority outcome.

At Walvoord's recommendation, CAC updated the spreadsheet of 10 UIW Core Learning Outcomes, and observed that, as of 2013, several outcomes had not had direct measures of student learning in the prior 10 years. CAC then looked at indirect measures on the longitudinal survey reports from NSSE and UIW's Graduation Exit Survey, specifically for those unmeasured outcomes. Noting a weakness in student's experiences in international topics, and given the increasing UIW priority in trying to provide more international experience, CAC selected Global & Historical Consciousness (G&HC) as the next outcome to measure. The latest outcome statement reads:

“...[students will be able to analyze] the ways humans have been influenced by cultural, religious, and ethnic traditions, by economic forces, and by environmental opportunities and limitations.”

The use of rubrics to score student writing had proven very effective in the previous core assessment, so CAC favored using rubrics for G&HC as well. The next question about how to identify and gather student work to assess became simpler in the context of a SACSCOC pre-conference workshop given by Dr. Catherine Wehlburg (Dec, 2013, SACSCOC Annual Meeting, Atlanta). Wehlburg demonstrated how to use a rubric designed for a core learning outcome, and use it to measure that outcome on student work where that outcome was not an explicit course objective. With that approach in mind, CAC decided to assess G&HC by using an established VALUE rubric, and simply gathering existing final exams and projects from subjects where G&HC is a natural outcome, and see what level of student learning could be observed.

The rubric we started with

CAC began with the AACU VALUE rubric for Intercultural Competence as its baseline. The rubric allows scoring of four levels of achievement on six learning outcomes:

- Cultural self-awareness
- Knowledge of cultural frameworks
- Empathy skills
- Skills in verbal and nonverbal communication
- Curiosity
- Openness

Courses and assignments collected

CAC invited faculty from several courses and disciplines to contribute graded collections of final papers, exams, and projects that the instructors felt might have elements where the rubric could help score the level of student learning in G&HC. Ultimately 200 samples were collected from the following courses:

World Literature	68 samples
U.S. History I	47
World History II	80
American Literature	5 (an upper-division course)

G&HC assessment working group

Full-time and part-time faculty from those departments were invited to join Dr. James (assoc provost for IE) and Dr. Crane (associate dean of humanities, and History professor) to conduct the scoring during a week-long summer workshop. Participating from the faculty were:

Dr. Gil Hinojosa, Professor of History
Dr. Pat Gower, Professor of History, Chair
Lesli Hicks, part-time instructor, History
Mary Ferrer, part-time instructor, History

The working group met June 9-12th in the GB 120 conference room.

Day 1 included introductions to each other, to the project, to the rubric, and to the ExamSoft software we used to log and analyze our rubric scores. We practiced scoring one sample, in order to calibrate our grading, as a group, and improve inter-rater reliability. In the course of that discussion, we modified the rubric, to simplify the number of outcomes we score, and to clarify the guidance for each level of learning.

The guidance for the group was to score student samples during the week to answer:

DOES THIS RUBRIC
WITH ANY OF THESE ASSIGNMENTS
HELP US MEASURE STUDENT LEARNING
OF THIS G&HC OUTCOME?

Our practice session led us to pair up in three sets of partners who would grade an entire section of work. We agreed on the following steps to grade each section:

- STEP 1. UNDERSTAND THE ASSIGNMENT
ANTICIPATE WHICH RUBRIC OUTCOMES TO LOOK FOR
- STEP 2. GRADE ONE OR TWO SAMPLES TOGETHER
DISCUSS:
WHICH RUBRIC OUTCOMES SEEM TO APPLY
WHAT EVIDENCE OF LEARNING CAN YOU GLEAN;
WHAT SCORES WOULD YOU ASSIGN
- STEP 3. SPLIT YOUR STACK OF PAPERS
DO YOUR HALF; RECORD YOUR SCORES
SWAP HALVES
DO THE SECOND HALF; RECORD THOSE SCORES
- STEP 4. REGROUP WITH YOUR PARTNER – MAKE NOTES FOR THIS STACK:
WHICH RUBRIC OUTCOMES FIT THIS ASSIGNMENT?
WHAT GENERAL LEVELS OF SCORES DID STUDENTS GET?
DID YOUR WORK SUGGEST ANY NEW OUTCOMES?
COULD THIS ASSIGNMENT BE:
USED ACROSS AN ENTIRE COURSE?
ADAPTED TO OTHER COURSES?

In Days 2 & 3 we graded all the student samples we had collected. Each of the 200 samples was graded twice, and every rubric was entered into ExamSoft to allow us to analyze the results afterward.

On Day 4, we reviewed our overall observations, created an even simpler final rubric that we agreed could be used across many disciplines. We also drafted bullets for final recommendations. Dr. James agreed to conclude the data analysis and draft the final report.

Findings

At the end of our workshop, we found that our slightly-modified rubric did indeed allow us to measure G&HC learning on these assignments. In general, we were more often able to discern student learning of “knowledge” and “attitude” outcomes, since the assignments were not designed to elicit demonstrations of skills. Depending which outcomes a faculty member wanted to assess, we definitely confirmed the utility of this rubric to measure student learning in G&HC.

Also, as we expected, student levels of learning tended to be low, since we scored student samples from courses which typically include freshmen and sophomores, and these courses are where students are first introduced to many of these outcomes.

In a broad sense, the working group was very encouraged by this finding. It meant that, even with assignments that were not specifically designed to measure these outcomes, evidence of G&HC student learning was still evident. This fact alone provides evidence of department-level reflection on core learning outcomes that resulted in deliberate integration of foundational core content in these required core courses. Moreover, since we saw evidence of learning across multiple courses, sections and faculty members, our findings indicate that core learning is emphasized well in faculty hiring, department activities, and curriculum development.

We also concluded that a few modifications would make the rubric instrument more useful and more applicable to multiple disciplines. We added a column “0” to allow the scorer to note “no learning” was observed or when the outcome was “not applicable” for a particular assignment. This feature is important, as we recall we want the rubric to be useful on assignments where the *core* learning outcome might not be explicitly stated. We also simplified the number of outcomes to three: one for content knowledge; one for application skills; one for “approach” or attitude toward cultural diversity. We also simplified the instructions in each level for students to demonstrate, and reduced the number of levels to “minimal”, “progress”, and “full integration”.

Since we examined a number of courses and sections, and exams or projects not originally designed to measure this particular core outcome, the working group identified several questions and prompts that proved to be more effective at discerning G&HC student learning. For programs and courses that would like to deliberately measure these outcomes, here are examples of the most useful question formats we observed; we hope other faculty can generalize these prompts for use in their own courses:

- a) Dr. Lonchar (World Lit) recalls a list of works covered during the term, and prompts students to pick two which “best illustrate what it means to be human in a ‘global’ culture”. The prompt tells them to “demonstrate knowledge of the two works...[what they] have in common and what distinguishes each....” This prompt allowed us to assess *content knowledge* of G&HC and *empathy*.
- b) Dr. Lonchar (World Lit) again recalls a list of titles encountered during the term, and provides a dozen small photos. Students are prompted to select a distinct picture for each title and explain. With this exercise, we were able to discern student *content knowledge, empathy, communication, and curiosity*.
- c) Dr. Lonchar (World Lit) asks students to “choose the main female character from each of the following works and point out...similarities ...and the key differences....” Since the listed works all had historical contexts, this question allowed us to observe multiple G&HC outcomes.
- d) Dr. Nath (World History II) asks students to “write a one page response ...” to the presentations given by other student groups in class. Since all the presentations had historical prompts, we were able to discern multiple G&HC outcomes from these papers.

- e) Dr. Hector Perez (World Lit) prompts with a critique of a work the students studied, and asks students to contrast the critic's work. This was a powerful prompts for G&HC outcomes: "Achebe claimed that Conrad's 'Heart of Darkness' was a racist text...a major motivation for Achebe's own novel.... Compose a short essay in which you argue that Achebe's novel in fact provides a literary antidote to what he saw in Conrad's.... include specific references...that Achebe may have found offensive....discuss how well Achebe's novel provides another view of Africa or Africans."
- f) Dr. Tabit (World History II) asks students to "compare and contrast the responses of China and Japan to Western imperialist powers in the 19th century. Why do you think responses varied as much as they did?" By prompting students to describe others' perceptions, this let us discern all three G&HC outcomes of *content*, *skill* and *attitude*.
- g) A World History II prompt provides a poster with a family scene, with the corresponding caption with which it was originally published. Students are asked to describe what the image tells us now about Communist China under Mao Zedong.
- h) Another World History II prompt asks students to "explore the role of religion in nation building in India and Pakistan."

Data analysis

With all the rubrics entered in ExamSoft, we were able to analyze the data several ways. Unfortunately, in June 2014, the software had not yet fully implemented the ability to log scores from multiple raters on the same student sample, so we had to do some work-arounds to manage the data. In particular, we had to enter two scores each, for a single student artifact, as if we had used two separate rubrics (which we did not), one for each rater. This made the data a little clumsy to analyze, however, scores could be exported into MS Excel, which allowed us to get statistics on our rubric scores.

Inter-rater reliability. We compared entire sections of scores done by both partners for two reasons. First, of course, we needed to be sure that scoring partners were judging consistently. Second, these comparisons would help give us confidence in the rubric's ultimate use for diverse courses and faculty members.

We were pleased to find good inter-rater reliability for workshop participants. In a 24-student section of World History II, for instance, partner scores on all six outcomes for this section were identical or differed by only one for 95% of the scores; this is strong similarity. Most reliability comparisons were between 95% and 100%, with one slightly weaker at 89%, and one outlier where rater scores for the knowledge outcomes (only) had no correlation; those knowledge outcomes were not counted in assessing student learning.

Validity. We relied on the many years of VALUE rubric assessment for the validity of the instrument to genuinely measure student learning.

Discrimination. The discrimination for rubric scoring is a measure of whether those students who scored high (or low) on a particular outcome also tended to score high (or low) overall. That is, the discrimination index gives some indication whether a particular question or rubric outcome (row) helps distinguish between strong and weak students. For all the sections we scored, the discrimination index for each measured outcome (some were not measured, depending on the exam) was at least 0.33, where 0.4 is considered to be high. This gave us confidence in the accuracy of our statistics.

Statistics. Rubrics were scored using the original six criteria of the VALUE rubric. However, most of the exam samples were able to assess some *knowledge* outcomes but not others, some *skills* but not others. Therefore, we determined that statistics of the scores will be most meaningful if we combined the scores with the simpler rubric, with one *knowledge* outcome, one *skills*, and one *attitude*. We migrated the scores into this simpler format before calculating statistics.

As seen in the summary table below, the average of all scores varied from 1.65 to 1.95. There was a significant difference (at over 99% confidence) in History scores where *knowledge* was higher, and in World Literature scores where *skills* were higher.

	know	skill	attitude
AVERAGE OF ALL	1.95	1.87	1.65
AVERAGE WORLD LIT	1.80	1.96	1.66
AVERAGE HISTORY	2.05	1.80	1.63

The number of students who scored a “2” or higher (i.e., a level of “progress” or “full integration / mastery”) provides a useful benchmark to establish goals in courses. The next table shows the number and percent of students who scored 2 or 3 in each of the three outcomes (in a set of 163 validly graded rubrics). Consistent with the lower average scores in attitudes, the percent of students who scored 2 or 3 in *attitude* was under 50%, where two-thirds of all students were able to score 2 (or “progress”) in both *knowledge* and *skills* outcomes. These levels represent reasonable baseline levels of achievement to compare outcomes in more advanced courses, in future assessments.

	know	skill	attitude
Number at 2 or 3	105	104	80
Percent at 2 or 3	65%	64%	49%

They also suggest a simple criterion for acceptable levels of student learning in lower-division courses. For instance, if we count the students who achieved at least a 2 in *knowledge*, and a 2 in *skills*, and a 1 in *attitude*, we find in these graded samples that 104 out of 163, or 64%, of students reached or surpassed this threshold. We also recommend this baseline level of achievement to compare with later assessments in upper-division courses.

Overall, then, the working group concluded that the student samples show reliable evidence that students completing the freshman History and Literature courses demonstrate acceptable levels of “minimal” learning in the G&HC *attitude* outcome, and “progress” learning in the G&HC outcomes of *knowledge* and *skills*.

Final rubric and recommendations

With the more simplified final form of the G&HC rubric, the working group brainstormed departments and programs where it can be used for courses or entire programs to assess student learning. The working groups recommended the following departments consider the new rubric for their use:

English World Literature
Government and International Affairs
Business International Business
History
Sociology Cultural Studies
Philosophy (course-dependent)
Religious Studies World Religions
Communication Arts Bilingual Communications
Study Abroad