

**University of Utah - Office of General Education
Learning Outcomes Assessment – Spring 2014
Report to the General Education Curriculum Committee**

Summary and Findings: Critical Thinking and Written Communication

In the Spring of 2014, the Office of General Education (OGE) conducted an assessment of 2 of the 15 General Education Learning Outcomes: Critical Thinking and Written Communication. This document summarizes the details of this assessment, reports findings, discusses outcomes, and discusses process recommendations.

Assessment Framework and Design

This assessment was the first comprehensive attempt to study the achievement of the U's General Education Learning Outcomes using examples of student work from General Education classes as evidence.

Over the past four years academic departments have been asked, as part of their General Education designation applications or renewal applications, to indicate which of the learning outcomes their courses meet, with the understanding that at some point in the future OGE would be asking for evidence of the achievement of those outcomes for assessment purposes.

In the fall of 2013 a request was sent to the departments of all courses that had chosen either the Critical Thinking or Written Communication learning outcome for their course over the past two years. This request asked departments to ask the instructors of these courses to submit, through a web link, four examples of student work: one of high quality, two of average quality, and one of low quality. This distribution of assignments was requested so that the whole range of achievement in each course was represented in the analysis.

The assessment tool that was used to score the artifacts was the set of rubrics that were designed by the American Association of Colleges & Universities (AACU) for these outcomes. Each outcome rubric has five criteria that describe the outcome, and scores that can be assigned to each of those criteria, including:

- 1: Baseline achievement
- 2-3: Milestone Achievements
- 4: Capstone Level Achievement

Reviewers were also told to give artifacts a score of 0 if they thought there was no evidence of the achievement of the criterion or an "NA" if they thought that the application of this particular criterion was not appropriate for this artifact.

The General Education Curriculum Council (GECC) members served as the reviewers. The Senior Associate Dean and Assistant Dean of Undergraduate Studies trained the Council members on the use of the rubrics. Two council members were assigned to score each artifact.

Process Results

The departments of 75 courses submitted 305 artifacts to OGE: 133 for Written Communication and 172 for Critical Thinking. Artifacts were pre-screened by OGE to remove any identifying information about courses, instructors, and student names. A number of these artifacts were eliminated from use because they had grading marks on the documents or because there was identifying information on the document that could not be removed.

From the remaining artifacts, 120 were randomly selected for review – 60 for each outcome. This subset was selected so that each Council member was assigned to review eight artifacts and each artifact was reviewed by two Council members. Of the 120 artifacts assigned, 86 (43 for each outcome) were reviewed by the two assigned Council members by the end of the evaluation period. The scores for those 86 artifacts are reflected in the results tables below.

Interrater reliability (IRR) - IRR was calculated using Spearman rank correlation (because the rating scale in the rubrics is ordinal level). After removing NA scores from the analysis the Spearman rank correlation for Critical Thinking was .42 and for Written Communication was .34. These estimates are lower than generally acceptable for IRR. However, there are several reasons why these rates make sense and that we expect them to improve in the future.

To begin, this study represents the first time all of these 30 raters have used the instrument, and OGE expects that reliability will improve with use. In addition, the design used in the current study utilized multiple raters and almost never the same raters for more than one artifact. Studies reporting higher IRR rates (.70 and higher) tend to use only two or three raters who score a common subset of assignments. This method accomplishes two things: it allows for increased reliability of scorers but is also more efficient and cost-effective because all assignments are not read. The current study was more concerned with all individuals on the GECC getting a chance to use the rubrics and the review process so that they could provide informed feedback about the process. In the future, fewer reviewers may be used on a subset of assignments which will likely result in higher IRR.

Outcomes

Figure 1 shows the overall distribution of ratings across all the criteria of both learning outcomes rubrics. This figure shows that the distribution of scores was quite normal with Written Communication (mean=2.0) receiving slightly higher scores on average than Critical Thinking (mean=1.94), although this difference was not significant.

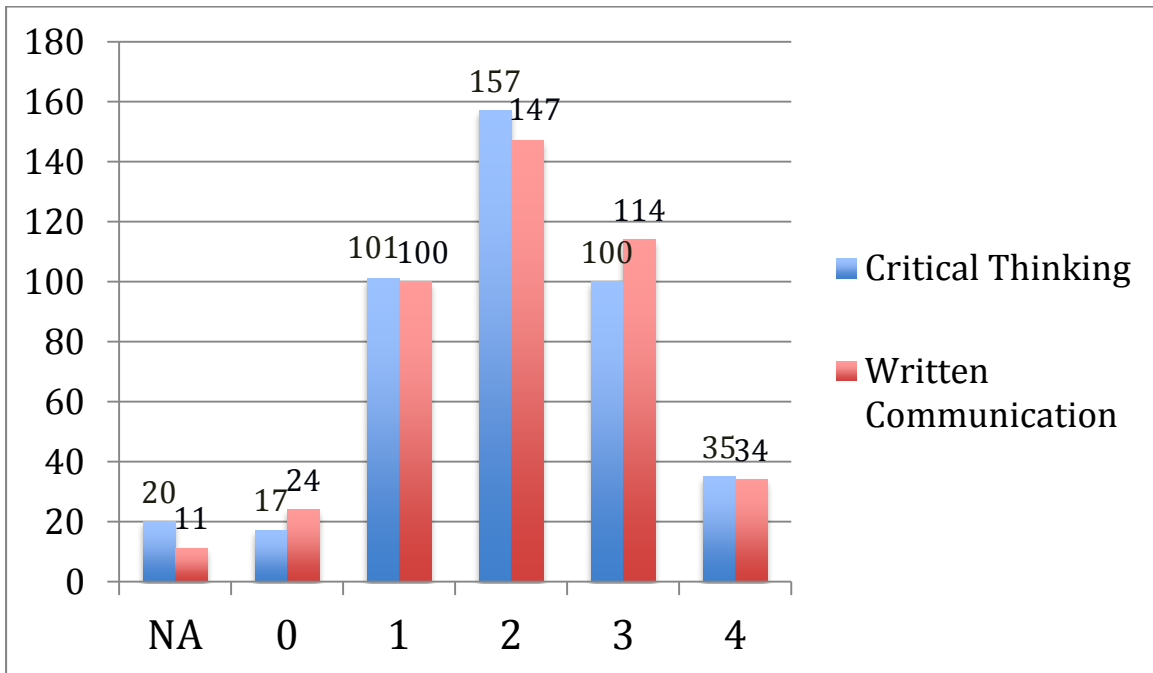


Figure 1: Overall Distribution of Learning Outcome Criteria Ratings: Written Communication and Critical Thinking

Figure 2 shows the distribution of ratings on two of the Critical Thinking criteria: Student's Position (perspective, thesis/hypothesis), and Influence of Context and Assumptions. The distribution of scores is quite normal and centers between 1 and 2.

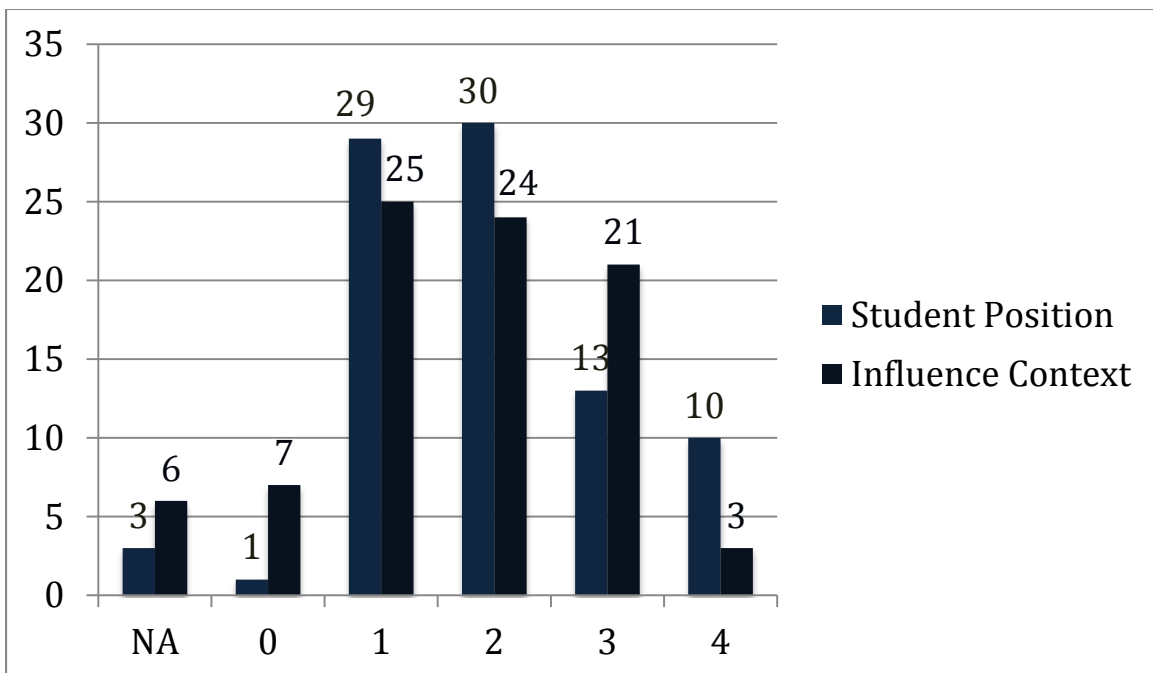


Figure 2: Distribution of Ratings on Critical Thinking Criteria of Student Position and Influence of Context

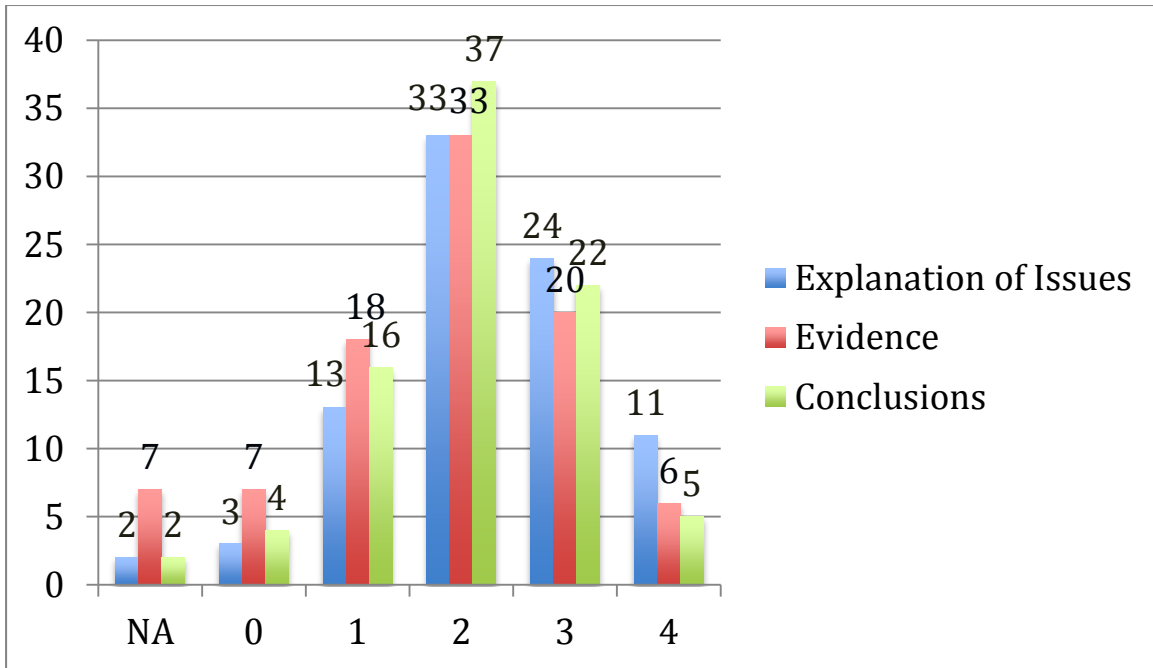


Figure 3: Distribution of Ratings on Critical Thinking Criteria of Explanation of Issues, Evidence, and Conclusions

Figure 4 shows the distribution of ratings for the Written Communication criteria of Syntax Mechanics and Sources of Evidence.

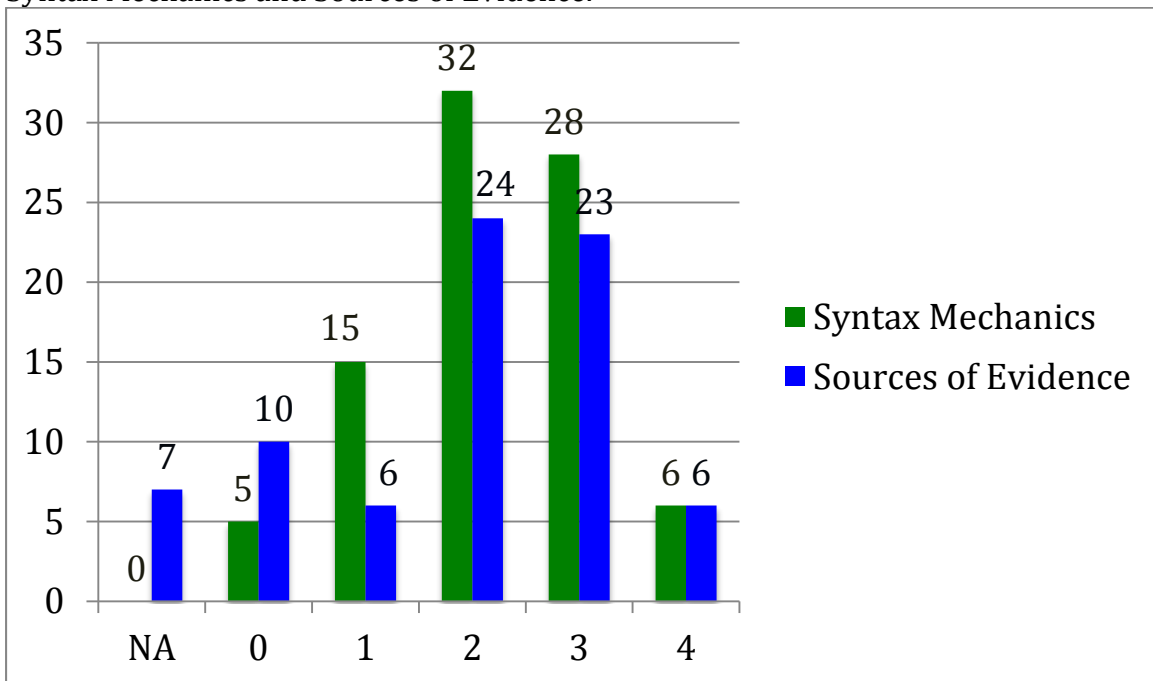


Figure 4: Distribution of Ratings on Written Communication Criteria of Control of Syntax Mechanics and Sources of Evidence

Figure 5 shows the distribution of ratings for the Written Communication criteria of Genre and Disciplinary Conventions, Context of and Purpose for Writing, and Content Development

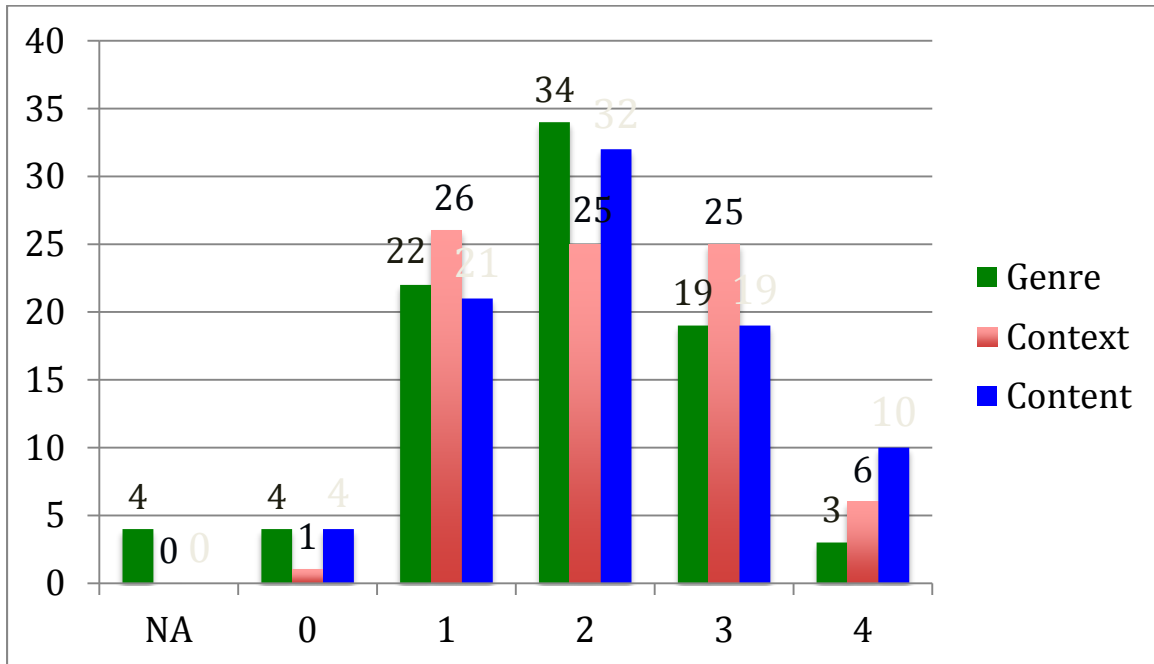


Figure 5: Distribution of Ratings on Written Communication Criteria of Genre and Disciplinary Conventions, Context of and Purpose for Writing, and Content Development

As mentioned above, most of the distributions of these rankings look fairly normal, with the average rating being about 2.0. On the AAC&U Rubrics, a rating of 1 indicates that students are at the baseline level of competence in the skill. A 2 is meant to indicate that they are at a first milestone of achievement in the trajectory of their college experience. It is not surprising to us that the students from the courses we received artifacts from, which were mostly 1000 and 2000 level courses, were centered around a 2, or a first milestone of achievement.

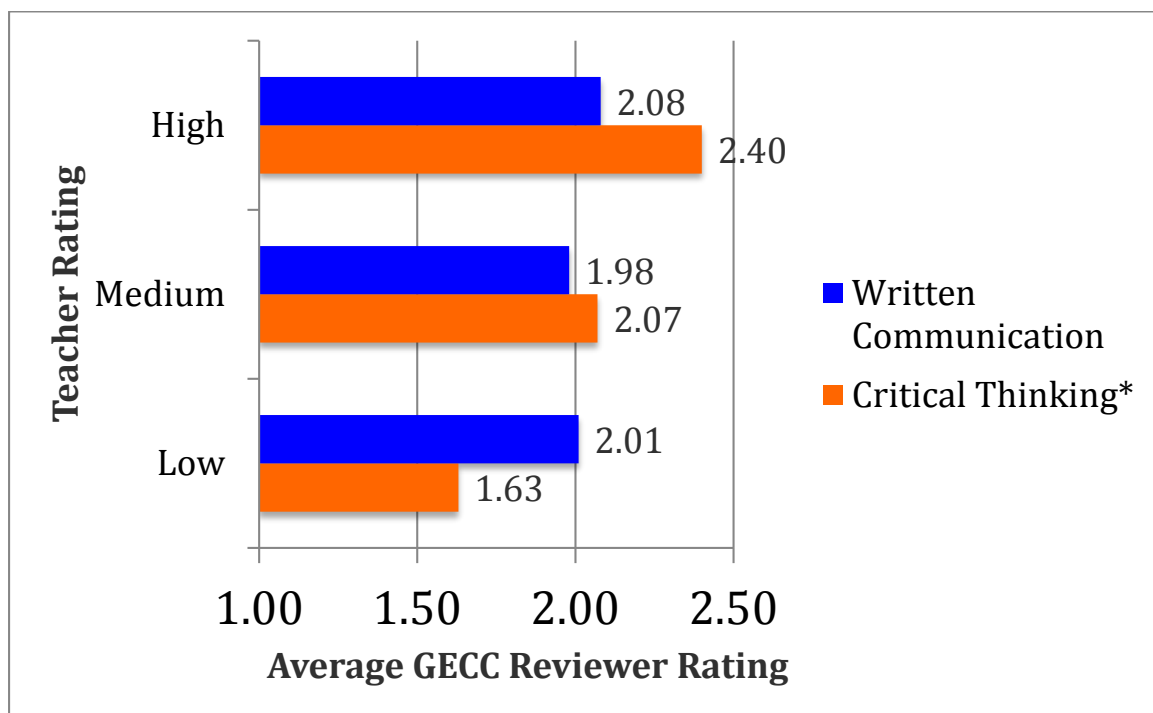
The achievement level described in these results is very similar to that which was observed in our 2013 General Education Learning Outcomes Assessment Report. Given that the current artifacts come from different courses within a year of those from the previous report, we did not expect to see very different results. Collecting these data on even larger data sets will help our office develop real baseline data to which we can compare future results. There are, however, some other ways to look at the data and the process to produce interesting and useful information.

Comparison of Teacher and Reviewer Ratings - In addition to observing the distribution of scores, OGE also examined the concurrent validity of reviewer ratings by comparing them to the overall quality category that instructors labeled their students' work with upon submission to the study. As a reminder, faculty were asked to submit one high, two medium, and one low quality example of student work for the selected outcome.

The overall correlation (again using the Spearman rank correlation) between teacher and reviewer ratings was .187, which is, again, quite low. However, when the analysis is done separately for Critical Thinking and Written Communication, differences arise. In Critical Thinking, the correlation is .339, which is statistically significant. However, there is virtually no relationship between teacher and reviewer rating for Written Communication (.026). See Figure 6.

One very likely reason for the lack of relationship between teacher and reviewer ratings for the Written Communication learning outcome was the absence of any artifacts that were submitted for the Upper Division Communication and Writing requirement. The distribution of designations for the courses from which the artifacts were submitted shows that the code for Upper Division Communication and Writing (CW) does not appear (see Figure 7).

Students in classes meeting the writing requirement would presumably be writing at a higher level, which would result in a greater variation in the scores, which would most likely result in a higher correlation between rubric scores and teacher ratings.



* $r=.339, p<.001$.

Figure 6: Spearman Rank Correlation Between Teacher and Reviewer Rating: Written Communication and Critical Thinking

GEDesig

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	AS	20	2.3	2.3	2.3
	BF	450	52.3	52.3	54.7
	DV	60	7.0	7.0	61.6
	FF	100	11.6	11.6	73.3
	HF	30	3.5	3.5	76.7
	IR	20	2.3	2.3	79.1
	SF	180	20.9	20.9	100.0
	Total	860	100.0	100.0	

Figure 7: General Education Designations of Courses from Which Artifacts were Submitted

PROCESS RESULTS

Because this was the first large scale assessment of learning outcomes in General Education, the GECC also discussed the process that was used to derive these data and observations. This discussion focused on three ways to improve the assessment process. These are discussed below.

Improve IRR. Assign teams of reviewers to the same artifact. Asking the same two reviewers to review the same artifacts of student work will improve the integrity of our assessments and enhance our IRR. Where possible, make reviewer teams based on subject matter alignment with the artifacts of student work. For example, rather than have an English Professor evaluate a piece of critical thinking from a nuclear engineering course, assign that writing to someone from Engineering.

Enhance Communication about Rubrics. The GECC observed that many of the student artifacts submitted to provide evidence of Critical Thinking were not well suited to review using the AAC&U rubric. Several recommendations were made about how to help faculty access and use the rubric to make decisions about which student artifacts to submit as evidence for the Critical Thinking learning outcome. Suggestions included giving the ELOs and VALUE Rubrics a more prominent placement on the UGS website, attaching the rubrics to our call for student artifacts, and offering faculty workshops about the rubrics. Each of these suggestions will be explored over the summer and targeted actions will be taken in the fall.

Strategize Sampling. Faculty were asked to submit four artifacts of student work, one representing low level work, two representing middle range work, and one representing high level work. The GECC discussed this sampling strategy at length. Ultimately, we would like to get to a place that would allow us to randomly sample a pool of archived student artifacts and we believe that increased use of Canvas will allow us to achieve that ultimate goal at some point in the future. In the meantime, the discussion focused on several options including asking for the top 20% of student work, or asking for only three samples of work, and looking into a purposeful stratified sampling technique.