**University of Utah General Education Learning Outcomes Assessment**

**Written Communication & Quantitative Literacy**

**Office of General Education**

**Summer 2015**

In the summer of 2015, the Office of General Education conducted an assessment of the Written Communication and Quantitative Literacy General Education learning outcomes (see ugs.utah.edu/gen-ed-reqs/outcomes.php to see all outcomes and their rubrics) using the American Association of Colleges and Universities’ (AAC&U) rubrics. This document summarizes that assessment.

**Purpose**

In addition to continuing to refine our process for our assessment of the General Education learning outcomes, another purpose of this study was to continue to investigate and improve our capacity to use the AAC&U rubrics to score artifacts (assignments, exams, etc.) generated by students taking General Education courses at the University of Utah.

**Participation in the Multistate Collaborative to Advance Learning Outcomes Assessment**

This review, and the selection of the Written Communication (WC) and Quantitative Literacy (QL) learning outcomes, coincided with a national project in nine states to study the assessment of the AAC&U’s Essential Learning Outcomes (ELO’s). This project was called the Multistate Collaborative to Advance Learning Outcomes Assessment (MSC), which was a joint project of the State Higher Education Executive Officers Association and AAC&U. These ELO’s are the same outcomes that were adopted by the University of Utah’s General Education program. The University of Utah participated in this national project along with the eight other state universities in the Utah System of Higher Education.

The following were requirements for participation in the MSC. The first was the selection of two of the three learning outcomes under study in the project: Written Communication (WC), Quantitative Literacy (QL), and Critical Thinking (CT). In 2014, we assessed WC and CT, but wanted to do WC again in order to get a wider range of artifacts to increase the validity of our assessment, so we selected WC and QL. The other requirement was to select ten classes meeting each of these requirements and collect ten artifacts from each of these classes from students who had completed at least 60 credits towards their degree total of 122 credits.

**Method**

*Artifact Selection* – The instructors of courses that selected one or both of these learning outcomes over the last four years during their General Education designation application or designation renewal were contacted. Courses that were likely to have students in them with 60 or more credits were targeted. These instructors were asked to produce ten assignments covering the range of achievement in their course. This resulted in 167 artifacts being collected from 17 courses. This was slightly less than what was requested for the MSC project, but met the needs nonetheless. 120 of these artifacts were used in the local analysis that is described in this document: 60 for each of the two outcomes. These 60 were chosen randomly within each learning outcome. Only 120 of the original 167 were analyzed based on the number of reviews we thought we could ask our reviewers to conduct.

*Rubrics for Analysis -* The rubrics were developed by AAC&U using a process that utilized faculty from institutions across the country. The University of Utah, in fact, sent a representative to one of AAC&U’s development seminars for the rubrics.

*Training* - Two faculty representatives from the University of Utah attended the national MSC conference, at which they were trained to use the AAC&U rubrics for WC and QL outcomes. These two individuals returned and trained 10 faculty members who are involved with General Education at the U. These 12 members were selected for this task, in part, for their expertise in the areas of WC or QL. Each of the review teams thus had 6 members. Each of these groups of 6 were split into 3 teams of 2. Each team reviewed 20 artifacts so 120 artifacts were used in the analysis.

**Analysis**

The unit of analysis for this study was the learning outcome across the whole Gen Ed program. As such, no identifying information was kept on students, courses or instructors. The only information about a course that was maintained was the level of the course (1000, 2000, etc.), and what college it was in and the Gen Ed designation that it met[[1]](#footnote-1). We also tracked what rating we asked the instructor to give to each artifact overall: low, medium, or high. This additional criterion was used as convergent evidence for assessing the construct validity of the rubric ratings.

**Results**

Included below are figures showing the distribution of competency levels on the criteria for each of the two learning outcomes under study: QL and WC. The rubrics from AAC&U include competency levels of 1 (“Benchmark”), 2 and 3 (“Milestones”), and 4 (“Capstone”). We added a “0” level to allow raters to indicate if they thought there was no evidence whatsoever of the achievement of the criterion. We also gave the raters the option to provide a rating of “Not Applicable” on a criterion (coded as “99” in the figures) if they thought the assignment, as designed, could not be used to rate the artifact in any way.

**Quantitative Literacy**

The distribution of ratings on the criteria for QL is in Figures 1 and 2 below. Those scores show a relatively normal distribution across the 0 to 4 range of achievement for each of the criteria. We would expect a normal distribution given that we asked faculty to produce 1 lower achieving student artifact, 2 medium achieving artifacts, and 1 high one. Seeing this distribution of scores is reassuring because it shows that the rubric is helping us identify the distribution that we asked for. This provides us with some evidence of validity for the use of the rubric for this purpose.

The other result that is reassuring is that the middle of the distribution of the scores is centered on the 2 and 3 milestone levels of achievement. The artifacts for the QL analysis came from mostly 2000- with some 3000-level courses. Given that these rubrics reflect a range of ability across the whole undergraduate experience, we would expect the distribution of scores to center on the middle of this scale and it was encouraging to see that.

It is also clear that there are criteria in the QL rubric that our raters felt could not be addressed by the kinds of assignments that were submitted. In particular, the scorers found it hard to use the Assumptions criterion (see Appendix A for the QL rubric) because the assignments did not necessarily ask students to describe assumptions, leaving raters without any way to judge it.

*Interrater Reliability*: Figure 3 shows two measures of interrater reliability between the two scorers for the QL artifacts. The red bars show the average real difference between ratings of the two reviewers, and the blue bars show the Spearman rank correlation (correlation for ordinal level data) between the two scores.

Both of these figures show that there is still quite a bit of discrepancy between the ratings of the two reviewers. In most studies of interrater reliability a desirable correlation is in the range of .70 or higher. The current levels hover just below .40, which is somewhat higher than in the previous study (mid to low .30’s), but it is still quite low. We attribute this low agreement to a couple of factors. First, the rubrics are still new to us, and this is the first time we have used the QL rubric. Also, half of the reviewers on the QL team were new to the General Education Learning Outcomes process, so this was essentially a new review for them. This level of agreement may persist until we either get very good at coaching reviewers on how to score artifacts and/or we maintain the same reviewers for future reviews so that we retain their expertise.

**Written Communication**

The artifacts for the WC outcome had a normal distribution centered at Milestone 3 for some of the criteria, and other criteria were negatively skewed, showing an increasing number of ratings as the graph moves toward the higher achievement categories. These artifacts showed a particularly good achievement on the Context and Sources-Evidence criteria.

This higher level of achievement (compared to previous review of WC and compared to QL in this study) was not surprising. Quite a few of artifacts came from courses at the 3000 level and higher, with a number of them coming from 5000 level courses that met the Upper Division Writing Requirement. One would expect those students to be performing toward the top end of this distribution of scores.

*Interrater Reliability* – Figure 6 shows the interrater reliability measures for WC. The Spearman rank correlation between raters was a little less than .40, like QL, which was lower than desired but still higher than the last time we assessed WC in Summer 2014. The average difference between the ratings of the reviewers was a little more than ¾ of a point.

**Discussion**

This pilot was successful in its overall purpose, which was to continue to introduce members of the General Education Curriculum Council to the assessment process we are using to examine the General Education Learning Outcomes, and to continue to refine that process as we learn from doing these assessments.

What we have learned in this pilot is:

* We should continue to do detailed trainings on the use of these rubrics.
* If at all possible, use some of the same reviewers for future assessments so we can maintain some institutional knowledge about how to use the rubrics this will improve the quality and consistency of assessment and our interrater reliability.
* We should work on making recommendations about the types of artifacts that lend themselves to assessment. We discovered in this assessment that the “Assumptions” criterion of the QL rubric was hard for reviewers to assess. We might consider removing that criterion, which would probably increase our interrater reliability, or make more specific recommendations about the type of assignments that we are seeking that would allow us to measure Assumptions.

1. At times, the identity of the course was impossible to de-identify based on the content of the artifacts provided. [↑](#footnote-ref-1)