**General Education Curriculum Committee**

**General Education Learning Outcomes Assessment**

**Intercultural Knowledge and Competence**

**Spring 2020**

**Introduction**

The General Education program at the University of Utah includes both lower division requirements that are state-mandated as well as upper division requirements that are specific to the University of Utah. The General Education program has 15 learning outcomes that were adopted in 2008 from the American Association of Colleges and Universities (AAC&U).

This report details the assessment of the Intercultural Knowledge and Competence (IKC) learning outcome in the Spring of 2020. This assessment was overseen and conducted by 12 members of the General Education Curriculum Committee (GECC) with the help of faculty who submitted artifacts from their courses.

**Method**

The IKC learning outcome was selected by faculty in 35 courses during a General Education five-year renewal or initial designation review application during the past four years (2015-2019).

**Recruitment**

The faculty of the 35 IKC courses were approached and asked to participate in this study by agreeing to do two things:

1. Use the Learning Outcomes Associator to make associations between the IKC learning outcome and an assignment in their course.
2. Allow the GECC to use the Learning Outcomes Reviewer to examine these artifacts with our rubric as part of our assessment.

After sending emails to all of the faculty of the 35 courses and conducting follow-up email correspondence, the following was accomplished:

* The instructors of ten courses agreed to participate.
* Two instructors were removed from the study because their assignment was going to be delivered too late in spring semester for it to be useful to this study.
* One course had to be eliminated because of technical problems with how the course was organized in Canvas that did not allow the Associator to access the IKC assignment (we are looking at solutions to this issue moving forward).
* This left us with seven courses (20% of the total 35 courses) in which the faculty agreed to participate in the study and complied with the requirements.
* Five of the seven courses were in the Humanities, one was in the School of Cultural and Social Transformation, and one was a course from the LEAP First Year Experience program in Undergraduate Studies.
* The seven courses included (interestingly) one 1000-level course, two 2000-level courses, and four 4000-level courses.
* In the seven courses, a total of eight different assignments were identified by faculty for inclusion in the study. This produced 200+ artifacts that were available to be sampled.

**Measurement Tool**

The GECC decided to use the AAC&U Global Learning rubric (see Appendix A) to evaluate this learning outcome because it better reflected how IKC was being taught by faculty at the institution. The existing IKC rubric had some problematic elements that were discussed during a IKC focus group a couple of years ago.

**Evaluation Method**

The 12 reviewers were split up into 6 teams of 2. Each team reviewed the same seven artifacts so each artifact had two raters. The Reviewer application was used to randomly select 42 artifacts from the total population of artifacts available (6teams x 7 artifacts).

**Results**

**Reviews**

One of the reviewers was excused from doing reviews because of unforeseen circumstances, but the rest of the reviewers completed their tasks.

The IKC rubric contains six ratings: Ratings 1-4 are labeled “Benchmark, Milestone 1, Milestone 2, and Capstone” describing increasing levels of accomplishment. A rating of “0” is also available, which indicates that there was “No Evidence” that the learning outcome was achieved. The rubric also includes an “NA” rating if reviewers believed that the assignment did not allow students to demonstrate the accomplishment of the criteria in the rubric.

Because of a temporary issue with how data from the Reviewer app were made available for analysis in the current study, it was impossible to distinguish between a rating of “NA” and “0”, so they are all combined into NA for the time being. When a .csv export is available, the ratings of “NA” and “0” will be separated.

Figure 1 shows the overall distribution of scores across all artifacts and gives an overall impression of the accomplishment of the learning outcome across the rubric. As mentioned earlier, the ratings in the rubric that aren’t “NA” or “0” range from 1 to 4, and with other AAC&U rubrics it is typical to think of those four levels of accomplishment as ranging across the student’s years at the institution. As students move through their education, one would expect their achievement level in the rubric to increase as well.

Figure 1 shows the mean score and 95% Confidence Interval for the IKC criteria. The mean for all criteria fell between 1 and 2. The lowest mean was for the “Application of Knowledge to Contemporary Situations” criterion and the highest was for “Global Self-Awareness.” The significant differences between criteria can be observed by noting which confidence intervals do not overlap with each other. Thus, the “Application of Knowledge to Contemporary Situations” criterion had a mean that was significantly lower than the “Perspective Taking” and “Global Self-Awareness” criterion.



Figure 2 shows the distribution of ratings across all criteria and all artifacts. This tells a similar story to Figure 1. The most frequent rating is a 1, with it being selected more than the other categories of 2-4 combined. The “NA” category, as discussed above, includes both the “0” and the “NA” ratings so it is difficult to interpret at this point. However, the existing NA rating does indicate that there were quite a few assignments that were not suited to the category and/or students who did not show any accomplishment on the criteria. This issue will be explored further when these ratings are separated.



**Interrater Reliability**

An important factor to consider in rubric rating studies such as this is the degree to which raters agreed on their ratings of the same artifacts. A simple intraclass correlation coefficient was used for this analysis. The reliability across all partners using a correlation coefficient was r=.32. This is a somewhat low interrater reliability. A score of .60 is a more acceptable level. However, this figure should improve when we split the “NA” ratings into “NA” and “0.”

Another way to examine interrater reliability in a rubric study is to break it down by the criterion. This allows the study to show if there was better agreement between raters on some of the criteria compared to others. Figure 3 shows the interrater reliability by criterion. These figures are also likely to increase once the NA ratings are corrected.

Figure 3 reveals that the reliability of ratings of “Perspective Taking” (r=.46) was more than twice that of the other criteria.



**Discussion**

The evaluation of the IKC artifacts proceeded as planned using the Associator and the Reviewer instruments. Although we had one rater not complete the reviews and one missing assignment description, the reviewers performed their tasks without significant incident. There was one image (poster) that could not be seen in its entirety by two reviewers, which impacted their ability to rate it. The issue of the “NA” and “0” ratings being combined for now is unfortunate, but can be rectified. Overall, the system for doing this assessment seems to be successful, with some small bugs to work out.

The overall ratings on the IKC rubric (see Figure 1) indicate that students are averaging between 1 and 2 on the 4-point scale. One might expect them to do better given that four of the courses were 4000-level, and only one was 1000. However, there was a significantly greater number of artifacts that were randomly chosen from the 1000 level course. As such, it is not surprising to see averages as low as 1.4 and 1.6. It is not currently possible to pull the students’ year in school from Canvas with the assignments, but adding this functionality would be helpful in terms of interpreting the performance on the rubrics. In the future, one way to rectify this issue would be to do purposive sampling across different levels of the courses so a more even distribution is achieved.

The interrater reliability of r=.32 is not at an adequate level. This is probably the result of using the Global Learning rubric to assess Intercultural Knowledge and Competence for the first time with a large group of people. Some of the 12 reviewers were more familiar with the rubric than others, which probably contributed to some of the inconsistency in the ratings. It should be a regular part of our process to norm the rubric with the group that is going to do the assessment. This will improve as more of the GECC become familiar with the new rubrics we will be using to assess our outcomes. The “Perspective Taking” criterion has low to moderate interrater reliability of .46 and we’ll see if that or any of the other reliabilities improves when we have the data cleaned up.

We asked the reviewers for their feedback on this assessment and the process used. They thought the Associator and the Reviewer worked very well to conduct the reviews. There were very few issues except a couple of assignments that two reviewers couldn’t see because of the way they displayed on a laptop. We will be addressing this with the developers.

Another issue that was raised by the reviewers was whether the Global Learning rubric itself was appropriate to the assignments. One reviewer said that she often had to put “NA” because she didn’t feel as though the assignments had much “global” content. Another reviewer said that the way the assignments were designed would never allow the reviewers to fully demonstrate all of the criteria of the rubric.

**Next Steps**

Moving forward, with the changes described above, this framework should be adequate for conducting assessments in General Education. Using the Associator and the Reviewer very significantly reduced the amount of time that we and the faculty providing us with artifacts spent on this study. This should help us get greater participation and, as a result, more valid results.

However, there is still the issue of the number of assignments that are getting “NA” votes. The GECC should work towards reducing this number as much as possible. One idea that has been suggested before is to hold workshops to coach faculty on the kinds of assignments that are needed in order for us to do a valid assessment with our rubrics. The more detail that faculty have on the rubrics, the better alignment there will be between assignments and the rubric, which will lead to a more valid assessment process.

**Appendix A**

[see attached Global Learning rubric]