

HLC General Education Assessment

The goal of this assessment process is to evaluate student writing in two domains, critical thinking and writing ability. Two rubrics have been adopted to be the basis for these assessments. Student writing samples are to be course assignments taken from general education classes with each sample to be scored by two or more raters. The purpose of this document is to provide some perspectives on developing the process.

I. Reliability

Scoring consistency is of major importance in a rating process. To enhance consistency some the following issues should be considered.

Interrater reliability: Each writing sample should be independently rated by two or more evaluators. The goal is to have evaluators reach an acceptable level of agreement when scoring.

Exemplars: The rubrics have three levels of scoring each with two sublevels resulting in a six point scale. Writing samples from participating professors that clearly reflect each of the three levels need to be gathered. Samples of high, medium and low writing ability, and of critical thinking, can be nominated by professors to form a pool of student writing examples. From this pool evaluators should work together to:

1. Identify samples that clearly reflect each of the three scoring levels
2. Identify samples within each scoring level that reflect the sub-levels
3. Highlight the language within each level and sublevel that was critical in placement within category by the evaluators. The exemplars should form a common basis for training and consensus reaching by evaluators.

Consistency: The training goal is for evaluators to reach an acceptable level of consistency in rating. Some variation in scoring will occur. The questions of how much difference between evaluators is acceptable and how the differences are to be managed will need to be agreed upon. One standard is for two independent evaluators to score document rating criteria to within one point of each other on the rating scale, for example a 3 and 4. An average of student performance can be determined if desired. If the difference between evaluator ratings is larger than one unit, for example a 3 and 5, a consensus process can be used during which evaluators reach an agreement concerning the rating. These differences can be used as a basis for discussions concerning further refinement of scoring.

Intra-rater reliability: Given the same information an evaluator should be consistent in their own rating when scoring at two different times. For practice and consistency,

especially if there is a significant time lapse between data collections, the exemplars provide a means for evaluators to reestablish their scoring standard.

II. Constructs

The two rubrics represent two different constructs being assessed, writing and critical thinking. It is desirable to treat them as independent constructs when scoring however this is difficult since the measure of critical thinking is, in part, a function of students' ability to write. There is a potential built-in scoring bias since good writing mechanics facilitate the expression of critical thinking. However, students who write well in terms of grammar and structure may reflect relatively little critical thinking. Other students may be thinking quite critically about a topic but unable to reflect their thinking clearly on paper. To ameliorate the problem two evaluators could alternate the scoring sequence by going through the writing samples twice. For the first time thru the writing samples one evaluator could score only the writing and the other evaluator only the critical thinking. Then, switch scoring assessments for a second time thru the sample.

III. Data Collection

Technology: It is desirable that the data collection method be somewhat standard. Do not mix paper-and-pencil collection with computer if possible. Mixing the format of collection adds another level of error in the process since facility with technology will be confounded with writing ability.

Sampling: Any student's ability to write and think critically will vary to some degree by subject area. If a large number of students are to be sampled, it is reasonable to gather one sample from each student as long as the sample covers the range of general education topics. If a smaller sample is planned, it would be desirable to have two or more writing samples from various subject areas for each selected student.

Length of writing sample: Writing samples are expected to be in the range of 350 to 700 words. While this may be adequate for scoring writing, I believe it may be a challenge to determine critical thinking skills within the same document length. For example, the highest rating level of Critical Thinking for "Considers context and assumptions" may be a fairly lengthy piece in of itself. To fit within this expectation, writing guidelines for students will need to be clear. However, the more precise the writing guidelines, the greater the potential for formulaic writing and the less likely writing will reflect an 'individual' component implied by rating criteria such as "Communicates own perspective..." in the CriticalThinking rubric.