

Assessing Writing in GE (Fosen, Blackstone, Trechter and Loker)

Assessment of writing in GE was guided by the overall GE Program Assessment process described in Part II, Background, above. GE Writing Assessment focused on direct assessment of student work already being produced for purposes of the courses involved.

The main steps in initiating the assessment included:

1. Task Force members contacted selected faculty teaching GE courses in spring 2006 and asked them to volunteer their course as sites for assessing GE writing. A conscious attempt was made to select courses for assessment from a variety of GE areas, including both upper and lower division courses. Assessing writing in GE was facilitated by the fact that a writing assignment of at least 1500 words is required in all GE courses.
2. Task Force members derived Student Learning Outcomes (SLOs) for writing in GE from [EM 99-05](#). The process was facilitated by prior work done by English composition faculty to revise EM 99-05 goals for Area A-2 in the direction of assessable outcomes, and by previous assessments of the freshman composition course, ENGL 130. GE Writing SLOs were further discussed and refined in discussion with faculty teaching GE courses. (See Appendix B for GE Writing SLOs.)
3. Task Force members, in consultation with faculty, developed a [GE Writing rubric](#) to assess student writing. (See Appendix B for GE Writing rubric.) To keep the assessment manageable, the Task Force made a conscious decision to limit both the scope and scale of the rubric. Written work was assessed along three characteristics (Content, Organization and Argumentation, and Grammar and Surface Features) and sorted into three categories (Beginning, Competent, and Accomplished) using criteria specified in the rubric.
4. Faculty who agreed to offer their courses as sites for assessment were asked to provide a course syllabus including a description of the writing assignment to be assessed.

During the steps outlined above, the Task Force had to make decisions regarding how many pieces of student work to examine, what sampling strategy would be followed for selecting student work, and what processes to use for recruiting and training readers. After several meetings of the Task Force, it quickly became evident that the assessment should offer a comprehensive and current “snapshot” of writing in GE, and that doing this necessitated the selecting a broad coverage of courses and reading a significant amount of student work. Because each piece of student work would be read by two independent readers, the Task Force faced a potentially overwhelming paper management problem.

To assist in the management of this process, the Task Force turned to an automated system for the storage, retrieval, assessment and analysis of student work, the [STEPS](#)© program. STEPS allowed students to upload their work directly to a central server where faculty readers could then access their assigned papers from any site with an internet connection. The great advantage of this system was the streamlining of the assessment

process, skipping entire stages of data input and “cleaning.” As readers completed their reading of student work, the results of their assessment were input directly into STEPS for subsequent analysis. Because readers would be assessing papers during busy weeks at the end of semester, their ability to read and rate papers at odd hours and for snippets of time was a boon to the assessment.

The downside of the use of an automated system such as STEPS was the need to train faculty in its use and the loss of the valuable conversations that can shape large-scale assessments. The Task Force faced the potential loss of the informal interchange of raters’ developing ideas and impressions of student work, and of the chance to reconcile divergent assessments of essays on the spot. Such informal interaction is an important aspect of program-level assessment because the joint, collective, and collegial reading of student work in a face-to-face environment can spur revision to the assessment instruments and immediate reflection on assessment methods and goals. Given the tight timeline for completing the assessment, as well as the huge gains in efficiency of an automated system, the qualitative benefits of face-to-face assessment were sacrificed in order to gain broader coverage of student work within the given time and budgetary constraints.

The next stages in implementing GE writing assessment necessitated the recruitment and training of readers of student work. From the outset the Task Force anticipated (a) need for two readings of from 400 to as many as 1000 pieces of student writing, depending on the rate of student participation in each course. Thus we anticipated requiring approximately 30 readers, each assessing up to 40 pieces of student work. Several of the faculty involved in earlier steps in the process (defining SLOs, creating the rubric, providing access to their courses) volunteered to be readers. The Task Force put out a call to all faculty and lecturers to recruit readers. Experienced graduate students were also recruited to fill out the ranks of readers.

The Task Force organized three mandatory two-hour “calibration workshops” to introduce readers to STEPS and orient them to the rubric with samples of actual student writing. The writings and their assessment were discussed openly in order to create consensus standards of what constituted work in the “1-Beginning,” “2-Competent,” and “3-Accomplished” levels of student writing. Readers were also trained in the use of the STEPS system at these workshops. Largely unfamiliar with STEPS, the readers reacted very positively to the software and felt comfortable using it to rate student work. Once readers were trained and ready to assess, an attempt was made to match readers with the content areas with which they felt most comfortable (humanities, social sciences, sciences, etc.). Readers received a stipend of \$125 for their participation in the reading/assessment process.

Results

Table 1 lists the types of courses that participated in the GE Writing assessment and basic information about the writing assignments that raters assessed. In order to allow students to acclimate to the content and disciplinary language of their courses and familiarize themselves with the grading procedures of their professors, we asked participating

teachers to have students upload papers written during the second half of the semester. We felt that essays completed later in the semester would be more representative of student ability than early ones.

On the whole, papers ranged in value from four to 25 percent of the overall course grade, and in length from two to ten pages. Based on information gathered from course syllabi and assignment sheets, about half of the papers seemed to be the first formal essays students had written for that class. Papers varied in content and organization depending on disciplinary conventions and individual goals of the instructors, but did fall into several broad categories. In lower division classes, the majority of papers were relatively low-stakes assignments and “think pieces” in which students responded to readings and for which first drafts and revisions were not required. A small percentage of lower division papers were formal analyses, research papers, and “case reports” that were revised and resubmitted based on peer or teacher response, and/or part of a structured sequence of writing assignments that helped shape the course. In upper division classes paper assignments generally followed conventions and organization strategies typical of writing in those disciplines; one example is the highly structured “case report,” which asked students to analyze a series of documents and develop policy proposals based on course concepts. Formal essays at the upper division were far more likely to be part of a structured assignment sequence involving in-class peer workshops, oral presentations, and multiple drafts.

**Table 1. Overview of GE Writing Assignments
Reviewed for Assessment**

Type of GE Course	# of papers assessed	# of paper for class	Week Paper Due	Paper Value (%)	Paper length (pp.)	Paper Genre
<i>Lower Division</i>						
Area A-2	66	2 nd	11	20	5	Rhetorical analysis
Area B-2	16	1 st	12	20	3	Case report
Area B-2	65	3 rd	10	4	2	Response
Area C-3	24	1 st	10	5	2-3	Think piece
Area C-3	48	3 rd	12	17	3-5	Summary/analysis
Area C-3	38	2 nd	8	10	3	Summary/analysis
Area C-3	39	1 st	8	25	5-6	Summary/analysis
Area D-3	19	1 st	12	25	10	Research/Analysis
Area E	76	2 nd	14	7	2-4	Think piece
<i>Upper Division</i>						
Area B	58	1 st	7	15	4-6	Case report
Area D	13	1 st	Varied	21	7-10	Research paper
Area D	69	1 st	11	8	1-2	Flyer

At the end of June, when essay rating was completed, the Task Force began the work of extracting analytical meaning from the data. The Task Force originally planned that they would serve as third readers to resolve any differences between the first two ratings. But the number of essays and the potential for two readers who disagree on even one criterion to trigger a third read made such a plan impractical. So we decided to average the score of the two raters. This left us with five potential scores (1, 1.5, 2, 2.5, 3) on any of the three dimensions of writing that were analyzed (Content, Organization, Grammar).

Table 2 presents a summary of the overall distribution of scores. There were slightly over 500 assessable examples of student writing that received two independent ratings for Content, Organization and Grammar. About 200 of these writing samples came from first-year students, 125 from second year students, 95 from third years and 85 from seniors.

Table 2: Frequencies of Ratings, by GE Course Division and First-Last Student Year

Criteria	Overall (%)	Lower Division (%)	Upper Division (%)	FY Students (%)	Seniors (%)
Content					
<i>1-Beginning</i>	51 (10%)	48 (13%)	3 (2%)	31 (15%)	4 (5%)
<i>1.5</i>	110 (22%)	89 (24%)	21 (15%)	51 (25%)	13 (15%)
<i>2-Competent</i>	186 (37%)	128 (35%)	58 (42%)	68 (34%)	34 (40%)
<i>2.5</i>	114 (23%)	70 (19%)	44 (32%)	39 (19%)	18 (21%)
<i>3-Accomplished</i>	46 (9%)	34 (9%)	12 (9%)	13 (6%)	16 (19%)
Total	507	369	138	202	85
Organization					
<i>1-Beginning</i>	61 (12%)	46 (12%)	15 (11%)	26 (13%)	8 (9%)
<i>1.5</i>	151 (29%)	125 (33%)	26 (19%)	77 (37%)	17 (20%)
<i>2-Competent</i>	150 (29%)	100 (27%)	50 (36%)	53 (26%)	27 (32%)
<i>2.5</i>	105 (20%)	70 (19%)	35 (25%)	35 (17%)	21 (25%)
<i>3-Accomplished</i>	47 (9%)	35 (9%)	12 (9%)	15 (7%)	11 (13%)
Total	514	376	138	206	84
Grammar and Surf. Features					
<i>1-Beginning</i>	83 (16%)	63 (16%)	20 (16%)	39 (19%)	11 (13%)
<i>1.5</i>	149 (29%)	115 (30%)	34 (27%)	67 (33%)	20 (24%)
<i>2-Competent</i>	171 (33%)	124 (32%)	47 (37%)	60 (29%)	33 (39%)
<i>2.5</i>	86 (17%)	68 (18%)	18 (14%)	33 (16%)	17 (20%)
<i>3-Accomplished</i>	24 (5%)	17 (4%)	7 (6%)	6 (3%)	3 (4%)
Total	513	387	126	205	84

About 390 pieces of student writing came from lower-division courses and 130 from upper-division GE courses. As can be seen in the table, 68% of all essays scored competent (“2”) or higher on Content, 58% scored competent or higher on Organization, and 55% scored competent or higher on grammar. The respective numbers for first-year students are 59% for Content, 50% for Organization and Grammar and Surface Features. For seniors these figures are 70% for Content, 71% for Organization and 63% for Grammar and Surface Features. There appears to be a slight but notable upward trend in the scores over time, a proposition that will be examined in more detail below.

As one can see above, a substantial percentage of scores in each of the three criteria reflected a “split decision” by two raters. That is, a first reader rated a paper a “1” and the second a “2,” resulting in a 1.5 average; the first a “2” and second a “3,” resulting in 2.5; or in some cases the first a “1” and the second a “3,” for an average of 2. Disagreement was higher in Content and Organization, up to 40% of all scores, than it was in Grammar/Surface Features. While reliability is usually lower in analytic as opposed to holistic scoring, there are two likely reasons for the low inter-rater reliability scores in this assessment. First, as discussed previously, STEPS allowed for efficient, cost-effective computer scoring of papers, but was not set up to re-create the valuable face-to-face conversations that enable raters to exchange ideas and reach a greater degree of consensus. Second, as there is no one standard for “good writing” across all disciplines in the university, the split scores seem indicative of authentic differences in the ways faculty understand and evaluate grammar, organization, and course content.

Scores were then analyzed statistically using basic descriptive statistics (mean, median, mode, standard deviation) and using t-tests and ANOVA to compare mean scores across groups. Descriptive statistics were generated for each of the three dimensions of writing examined for all of the essays as a whole, and then examined separately for first-year students versus seniors, upper- versus lower-division GE courses, transfer vs. “native” students and by the college of the student’s major (students from Natural Sciences versus Humanities versus Business, etc).

Content. The mean score on Content for all papers ($n = 507$) was 1.99 with the median and mode both equal to 2. (See Appendix D for summary of numerical data.) When comparing first-year students and seniors, the Content scores for first-year students was 1.88, with median and mode equal to 2, while the corresponding scores for seniors was 2.17, with median and mode equal to 2. The difference in mean scores was analyzed using an independent samples t-test and found to be significant ($t = 4.007$, $p < 0.00$). Thus, seniors score significantly higher than first-year students on Content. While the difference may seem small (1.88 versus 2.17), this increase of 0.29 points on a scale with a range of 2 points (1 to 3) represents a positive change of about 14.5%.

A similar trend was noted for lower- versus upper-division courses. Mean Content scores were higher for Upper Division courses (2.15 versus 1.94) and the difference was statistically significant ($t = 4.35$, $p < 0.00$). There was no significant change in Content scores between “native” versus transfer students, nor were significant differences detected by college of major. Thus the major trends in Content scores were that seniors

scored significantly higher than first-year students, and papers in upper-division courses scored significantly higher than lower-division courses.

Organization and Argumentation. Organization scores followed a trend similar to those of Content. The overall mean Organization score for all papers ($n = 514$) was 1.93, with the median equal to 2 and the mode equal to 1.5. (See Appendix D.) In the mean Organization scores of first-year students (1.84) versus seniors (2.06), the difference proved statistically significant ($t = 2.93$, $p < 0.00$). As with Content, a similar trend was noted for upper versus lower-division courses: the mean Organization score for lower-division courses (1.90) was significantly lower than that for upper division courses (2.01; $t = 1.99$, $p < 0.05$). As with Content scores, there was no significant difference in Organization scores between “native” versus transfer students, nor were significant differences detected by college of major. Thus the major trends in Organization scores were similar to, but weaker than, those found in Content scores: a trend toward significantly higher scores for seniors versus first-year students and for papers in upper-versus lower-division courses.

Grammar and Surface Features. Trends in Grammar and Surface Features were similar to those of Content and Organization, though the trends were weaker. The overall mean Grammar score for all papers ($n = 513$) was 1.82, with the median and mode both equal to 2. (See Appendix C.) The mean Grammar scores of first-year students was 1.76 with the median and mode equal to 1.5. The mean Grammar scores of seniors was 1.89 with median and mode equal to 2. This difference in mean Grammar scores is statistically significant under the assumptions of a one-tailed test ($t = 1.91$ and a one-tailed probability of $p < 0.026$). Unlike Content and Organization, there was no significant difference in Grammar scores between lower- and upper-division courses. As with Content and Organization scores, there was no significant difference in Grammar scores between “native” versus transfer students, nor were significant differences detected by college of major. Thus the major trends in Grammar scores show some similarities to Content and Organization, but are more ambiguous overall. Under the rating scheme employed in this study, it appears that seniors have slightly stronger grammatical skills than first-year students, but the difference in skill level in the two groups is weaker than for Content and Organization.

In summary, the analysis of assessment scores reveals statistically significant trends in the expected direction: our seniors in GE courses tend to score higher on all measures of writing performance (Content, Organization and Grammar and Surface Features) than do our first-year students. Clearly a number of methodological issues can be raised about how the papers were collected and assessed, the techniques used in this analysis, and the interpretation of the results. But as a first “snapshot” of writing in GE on our campus, the results raise some interesting and useful questions about the quality of our students’ writing, how writing is taught in GE, how writing assignments are crafted, and our understanding of writing at CSU, Chico. While these results are heartening, the gains exhibited in writing do not merit complacency with writing instruction on our campus.

Conclusions

The GE Writing assessment process provided valuable insights into our students' writing ability, the practices of GE writing on our campus, and the assessment process itself. Each of these areas will be discussed in this concluding section.

Student Writing. Based on this sample of courses, writing would seem to be a frequent and important part of most General Education classes. The Task Force had no trouble finding lower- and upper-division courses that were using writing to support student learning and assigning tasks typical of college-level thought. The analysis of assessment scores reveals statistically significant trends in the expected and hoped for direction: seniors in GE courses tend to score higher on all measures of writing performance (Content, Organization and Grammar and Surface Features) than do our first-year students. Content scores increased 14%, and gains in Organization and Grammar and Surface Features are slightly smaller.

In the context of research on writing, these gains appear respectable but not ideal. As students move through college—through GE and into their majors—writing tasks increase substantially in terms of their cognitive complexity, the content-specific language they ask students to draw on, and formatting and layout requirements. With some notable exceptions, lower division students here (omit?) wrote short papers in which they responded to course readings using their own evolving opinions of course content, while upper division students were asked to synthesize multiple concepts and sources of information to inform readers of an issue, argue a point, or develop policy recommendations. Upper division assignments, more often than at the lower division, asked students to shape their writing for a public audience outside of or in addition to the teacher, and to research disciplinary and scholarly sources to persuade that audience. Under these circumstances, the scores of seniors and upper-division students seem indicative of larger gains in rhetorical, disciplinary, and critical thinking ability than can be seen in the numbers. At the same time, it seems telling that the differences in scores are not more pronounced. As we will discuss below, our conclusions about the writing itself are tempered by the limitations of this pilot assessment.

Writing pedagogy in GE. Because writing tasks vary based on audience, purpose, and the conventions of a genre, gathering the writing assignments and syllabi from each GE course we assessed was an essential aspect of analyzing student writing. These documents gave our readers a useful, but admittedly incomplete, idea of the classroom contexts out of which the papers emerged: the texts students were reading, the assignment plans, the overall schedule of the course, and the place of writing in it. If the syllabi and assignments we gathered are any indication of GE as a whole, it seems clear that there is a good deal of variety in the types of writing assigned in GE classes. Classes varied in terms of the number of papers students wrote, the week of the semester they were due, their value toward a final grade, their length, and the genre or form they took. Assignments also varied widely in terms of the degree of structure and direction provided by professors, the amount of integration of the writing with other class activities, and the use of multiple drafts, peer or teacher response, collaboration, and other process-oriented instruction.

It is crucial that faculty have the latitude to create writing assignments that best fit their goals for their courses, consistent with their disciplinary expertise, their student learning goals, and the guidelines for General Education. It is also important to recognize that faculty not specifically trained in writing instruction may have questions or concerns about using writing as a mode of learning, crafting activities to support writers, and responding to and grading student work. They may have questions about the utility of the current 1500-word writing requirement for courses in General Education, or how best to implement it in large-enrollment courses. Thus faculty may need support in crafting meaningful writing assignments that help students develop their writing skills and engage them in the intellectual habits of the university.

Writing Assessment. The GE Writing assessment process outlined here was an experimental effort carried out at an accelerated pace. Many lessons were learned that can inform future efforts. Task Force members feel that the results obtained provide valid information that can help inform writing instruction and practices on our campus and recommend that GE Writing assessment be institutionalized as part of ongoing assessment efforts. The process of assessing large quantities of student work was greatly facilitated by carrying out *embedded* assessment of work already assigned in the classes chosen for this effort. As such, participation in the assessment occasioned only a minor increase in work load for both faculty and students. Workload was also eased by the use of the STEPS automated assessment and tracking system that facilitated collecting student work, allocating work to evaluators, collecting and storing evaluator's assessment results and storing this data in an easily accessible format amenable to quantitative analysis.

The present assessment offers a snapshot of writing in General Education. A more comprehensive picture of student writing at CSU, Chico, should work to capture a more continuous view of writing over longer periods of time. Indeed, research confirms that multiple pieces of writing gathered over semesters or years offer far more complete and valid assessments of student writing ability and development. Portfolio assessments within one course, longitudinal studies of students moving through GE courses, and multi-methodological studies of qualitative and quantitative criteria in writing offer much promise in capturing these gains in more complex, authentic ways.

Recommendations

1. Results of this assessment should be widely shared on campus to encourage an in-depth discussion of student writing.
2. GE Writing assessment should be continued at regular intervals for the indefinite future in order to monitor student learning and progress, especially in response to curricular and pedagogical innovation.
3. Continued collection of student writing will enable true longitudinal study of writing development during General Education, the major, or a whole academic

career at CSU, Chico, providing a more authentic view of writing on our campus.

4. Future writing assessment efforts should:
 - a. strive for a more systematic and representative sampling of GE courses;
 - b. provide stronger, more sustained training for readers involved in assessing student writing through more intensive calibration workshops;
 - c. continue to use STEPS to manage this process efficiently and permit the storage of large samples of student work;
 - d. seek to recapture some of the informal communication among evaluators of student work to foster conversations about writing as well as reduce inter-rater discrepancies.
5. Generally speaking, faculty value writing as the *sine qua non* of an educated individual. Faculty need to align their pedagogical practices with these values and continue to seek out creative, effective ways to engage students in constructive, developmental writing practices.
6. The university needs to provide ongoing faculty development that supports efforts of faculty at all levels to craft effective, appropriate GE writing assignments that balance workload with best practices in writing instruction.
7. Future GE Writing assessment should strive to systematically examine the effects of varying writing activities on student learning.
8. The GE Writing SLOs and rubric should be widely circulated on campus to elicit further discussion and refinement of these tools, and shared with students to provide them with clearer expectations of writing in GE.