GE Assessment Report, 2005-2006

GE Assessment Task Force
Don Alger, Chair GEAC
Phyllis Fernlund, Chair AURA
William Loker, Dean of Undergraduate Studies

GE Writing Task Force
Chris Fosen, Chair
English, GEAC
Sarah Blackstone, Dean HFA, AURA
Sara Trechter, English

GE Oral Communication Task Force
Ruth Guzley, Chair
Communications, AURA
Mitchell Johns, GEAC
Susan Avanzino, Communications

GE Quantitative Reasoning Task Force
Margaret Owens, Chair
Associate Dean, CNS, AURA
Russ Mills, GEAC
Jack Ladwig, Mathematics

Executive Summary: GE Assessment, 2005-2006

AY 2005-06 saw a major effort to assess student learning outcomes (SLO) in General Education at California State University, Chico. This work involved the combined efforts of the General Education Advisory Committee (GEAC) and the All University Responsibility for Assessment (AURA) committee, as well as the participation of more than 75 faculty members across the university from all Colleges. GE program assessment is mandated in EM 99-05. The current effort represented a significant departure from past GE assessment efforts, striving to go beyond course assessment to examine foundational skills – oral communication, writing and quantitative reasoning – across a variety of GE courses. Task Forces composed of members of GEAC, AURA and additional faculty followed a common strategy of:

- using EM 99-05, in consultation with faculty, to define SLOs for these foundational skills,
- identifying courses in GE as sites for assessment;
- working with the faculty who teach these courses to identify assignments that could be used for embedded assessment of student learning on the SLOs identified,
- analyzing the results of student work and drawing conclusions about GE instruction and learning from this analysis.

The assessment process is discussed in more detail in Section II and Appendix A, with specific details on the assessment process for each of the domains examined in its respective portion in Section III.

Among our results and recommendations are the following:

1. A variety of GE writing assignments of appropriate complexity and academic rigor were identified in the courses examined. On the other hand, appropriate quantitative reasoning and oral communications tasks were not readily identified outside of courses specially designated for those purposes in Area A-1 and Area A-4. The lack of quantitative reasoning and oral communication tasks raises serious concerns, given the stated intent of campus GE policy: EM 99-05 states that "[i]n every course, relevant skills of the Core must be applied as essential to the process of mastering content and making applications." (emphasis added). This objective is further reinforced in EM 99-05's discussion of the capstone requirement, stating that "[t]hemes will incorporate, build upon, and nurture skills from Area A…" *Our report calls for a campus conversation on how best to incorporate quantitative reasoning and oral communication throughout GE, particularly in the Upper Division Themes.*

2. Cross-sectional data comparing the performance of first-year students versus seniors revealed statistically significant gains in writing in GE courses in the three dimensions of writing examined: Content, Organization and Argumentation, and Grammar and Other Surface Features. This same trend was not apparent in Oral Communication, where the examples of student oral performance in Area A-1 courses (CMST 131, 132) were, on average, of higher quality than presentations examined from Upper Division Theme courses. Cross-sectional data on quantitative reasoning did not reveal any significant difference in the performance of first-year versus other students on either of the two quantitative reasoning tasks assessed (a probability problem and a calculus problem). *Cross-sectional data reveal a mixed picture of our*

students' performance and development in the foundational skills examined. Gains, where present, were modest and some skills showed no gains or deterioration. We call for a campus-wide discussion of these results as well as ongoing assessment to better capture changes in student performance in these areas, as the cross-sectional data examined here have limitations in this regard.

3. The Writing Task Force noted great variability in writing assignments, as well as among our readers in assessing writing. The variability in assignments is to be expected given the variety of courses examined, faculty prerogatives in creating assignments, and is not, in itself, any cause for concern. The lack of practice and "nurturing" of oral communications and quantitative reasoning skills in GE, and the apparent lack of "progress" by more advanced students in these areas, is more alarming. *We recommend that the university provide ongoing faculty development that supports efforts of faculty at all levels to craft effective, appropriate writing, oral communications and quantitative reasoning assignments, within and outside GE, that balance workload considerations with best pedagogical practices in these domains.*

4. We feel that Task Force members, in consultation with faculty colleagues, made good progress in terms of defining SLOs in core areas of GE, as well as in creating rubrics helpful in assessing and grading student work. From this follows two recommendations: *(1) SLOs and rubrics should be widely circulated to faculty for further input and refinement so that these can be adopted and used, <u>where faculty so desire</u>, as well as shared with students to clarify expectations regarding student performance in these areas. (2) The elaboration of SLOs and rubrics should be extended to other domains of GE including critical thinking, and coursework in Areas B, C, D and E. The creation of shared SLOs in GE will help clarify our objectives in GE instruction and provide the basis for ongoing assessment of GE courses.*

5. Task Force members, and many of the faculty who worked with us on this project, found the process and results extremely interesting and instructive regarding the state of GE and student learning on our campus. *It is important that this process, or one similar, be continued and extended to other domains of GE, not as a substitute for course-by-course review, but as an important support and addition. This process needs to be kept as unobtrusive and efficient as possible to avoid overburdening faculty (and students).*

As a first step in assessing the GE program, we feel this effort is a success. It is also incomplete. *The next steps of disseminating results, discussing and applying these results to curricular and pedagogical change in GE are critically important*. Those who carried out this work are divided as to whether the activities outlined here represent true GE <u>program</u> assessment, or simply "(sub-) Area" assessment. We need to consider alternate strategies, as well as modifications to the GE program itself, to get at a clearer picture of the contribution of GE to our students' overall educational experience. We hope the issues raised in this report will advance this discussion.

II. Background: GE Assessment

The primary goal of the GE Program Assessment carried out in AY 05-06 is to assess the degree of success of General Education efforts on our campus by observing the level of student learning as exemplified by student performance on tasks directly related to GE learning goals. GE Program Assessment is expected to generate additional benefits by fostering discussion of General Education goals and outcomes on our campus and the sharing of assignments and pedagogical strategies used by faculty to reach these goals.

The MOU from Academic Affairs that governs the administration of GE on our campus mandates GE Program Assessment as a regular part of the administration of GE. Review of GE is the responsibility of GEAC; program review was scheduled for AY 04-05 as part of the regular cycle of course reviews. In AY 04-05, GEAC sought faculty feedback on the course review process as part of program assessment. The AURA committee is charged with supporting the assessment of student learning outcomes (SLOs), including GE. Established in 2003, AURA has been working with departments on establishing procedures for assessment of SLOs in baccalaureate programs across campus. At the end of AY 04-05, the committee turned its attention to assessment of SLOs in GE.

The structures and processes for GE Program Assessment were worked out at a series of meetings between members of GEAC and AURA in the summer of 2005. These meetings resulted in a series of agreements outlined in a document specifying joint GEAC and AURA responsibility, a structure and process for carrying out GE Program Assessment, and a timeline for completion of the tasks outlined (See Appendix A). Members of the two committees agreed to share the work GE assessment, with an initial pilot project focused on the assessment of SLOs in core skill areas of GE, including Writing, Oral Communication and Quantitative Reasoning. Members of the committees agreed that assessment would:

- primarily involve direct assessment of student work assigned as part of GE courses to avoid over-burdening faculty;
- focus on student learning outcomes (SLOs) derived from relevant documents (EO-595, EM 99-05, MOU) shaping GE;
- include broad consultation with faculty on all aspects of GE assessment including definition of SLOs, the identification of relevant student work and the actual process of assessing student work.

The structure we put in place created a Coordinating Committee comprised of the Chairs of GEAC and AURA plus the Dean of Undergraduate Studies. Three task forces were created, focused on each of the skill areas chosen for assessment (writing, oral communications, quantitative reasoning). Each task force included a member of GEAC, a member of AURA plus an additional faculty subject matter expert. Additional faculty teaching relevant GE courses were recruited by each of the task forces to participate in the assessment. In all, over 75 faculty from across campus in participated in GE assessment activities with participation varying from offering their courses as sites for assessment, helping to define SLOs, creation and critique of rubrics used to assess student work, actual assessment of student work and analysis and write-up of results. All activities reported here were carried out from June 2005-July 2006.
GE Program Assessment in each of the domains chosen, Writing, Quantitative Reasoning and Oral Communications, was guided by the same principles: direct, embedded assessment of

student work based on SLOs derived from relevant GE documents. SLOs were discussed with faculty stakeholders and student work was assessed using rubrics created with faculty input, in classes voluntarily offered for assessment by faculty. Student work was assessed by faculty members who offered their services and were compensated for this work. Despite the common framework for carrying out the assessment, each domain examined presented special challenges and circumstances that necessitated creatively adapting the general model to particular circumstances. The specific procedures followed are outlined in the sections that follow for each domain assessed.

III. Assessment Results

A. 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
Lower Division						
Area A-2	66	2nd	11	20	5	Rhetorical analysis
Area B-2	16	1st	12	20	3	Case report
Area B-2	65	3rd	10	4	2	Response
Area C-3	24	1st	10	5	2-3	Think piece
Area C-3	48	3rd	12	17	3-5	Summary/ analysis
Area C-3	38	2nd	8	10	3	Summary/ analysis
Area C-3	39	1st	8	25	5-6	Summary/ analysis
Area D-3	19	1st	12	25	10	Research/ Analysis
Area E	76	2nd	14	7	2-4	Think piece
Upper Division						
Area B	58	1st	7	15	4-6	Case report
Area D	13	1st	Varied	21	7-10	Research paper
Area D	69	1st	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					
1-Beginning	51 (10%)	48 (13%)	3 (2%)	31 (15%)	4 (5%)
1.5	110 (22%)	89 (24%)	21 (15%)	51 (25%)	13 (15%)
2-Competent	186 (37%)	128 (35%)	58 (42%)	68 (34%)	34 (40%)
2.5	114 (23%)	70 (19%)	44 (32%)	39 (19%)	18 (21%)
3-Accomplished	46 (9%)	34 (9%)	12 (9%)	13 (6%)	16 (19%)
Total	**507**	**369**	**138**	**202**	**85**
Organization					
1-Beginning	61 (12%)	46 (12%)	15 (11%)	26 (13%)	8 (9%)
1.5	151 (29%)	125 (33%)	26 (19%)	77 (37%)	17 (20%)
2-Competent	150 (29%)	100 (27%)	50 (36%)	53 (26%)	27 (32%)
2.5	105 (20%)	70 (19%)	35 (25%)	35 (17%)	21 (25%)
3-Accomplished	47 (9%)	35 (9%)	12 (9%)	15 (7%)	11 (13%)
Total	**514**	**376**	**138**	**206**	**84**
Grammar and Surf. Features					
1-Beginning	83 (16%)	63 (16%)	20 (16%)	39 (19%)	11 (13%)
1.5	149 (29%)	115 (30%)	34 (27%)	67 (33%)	20 (24%)
2-Competent	171 (33%)	124 (32%)	47 (37%)	60 (29%)	33 (39%)
2.5	86 (17%)	68 (18%)	18 (14%)	33 (16%)	17 (20%)
3-Accomplished	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 abut 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.

B. Assessing Quantitative Reasoning in GE (Owens, Ladwig, and Mills)

Introduction

Many students at CSU, Chico, receive much of their college-level mathematics education from the one MATH course they complete to satisfy the GE Core A4 requirement. Students who receive additional, significant exposure to college-level mathematical principles and applications are typically enrolled in academic programs having some additional quantitative requirements.

EM 99-05 states that "[i]n every course, relevant skills of the Core must be applied as essential to the process of mastering content and making applications." This objective is further reinforced in the capstone requirement, which dictates that "[t]hemes will incorporate, build upon, and nurture skills from Area A…" Since mathematics is part of the Core (i.e., Area A), this requirement strongly implies that mathematical content should be distributed throughout the GE program, not just in MATH courses and a few science courses. However, this implication has not been systematically evaluated during any previous assessment of the GE program nor has it been encouraged or enforced other than within Sub-Area A4.

Other academic areas housed in the GE Core – written communication, oral communication, and critical thinking – have to varying degrees received greater attention to interdisciplinary application than has mathematics. This can most readily be explained by the relative newness of mathematics to the Core. Previously, the mathematics component of GE was grouped with the science requirement, as is the case with the statewide description of general education. With the adoption of EM 99-05, mathematics at CSU, Chico has now been aligned as a core area in order to recognize the central nature of mathematics in general education. That is, mathematics is not a culminating experience. Rather, it is a basic yet essential component of knowledge having potential to facilitate understanding and ability in a breadth of academic subjects. Indeed, mathematics forms the foundation for quantitative reasoning, which can be defined as the application of mathematics to describe, analyze, and solve authentic problems in context.

Mathematics in general education at CSU, Chico has previously been assessed in two different ways. The most significant and continuous assessment means has consisted of regularly scheduled GE course reviews. Area A, including Sub-Area A4 Mathematics, was last reviewed during academic year 1999-2000. No significant deficiencies in Sub-Area A4 Mathematics were discovered during the last review. Additionally, during AY 2001-02, a pilot assessment project was conducted, focusing on one course in the GE program, MATH 005, Statistics. The Learning Environment Assessment Project – Mathematics (LEAP-M) involved both direct and indirect measures of student learning. Lessons learned in the LEAP-M project proved valuable in designing the probability component of the current project.

Process

1. Planning the Process

a. Develop a list of Area A4 GE student learning outcomes.

In the late 1990s, a CSU-sponsored project solicited input from three campuses, CSU, Chico, CSU, Sacramento, and San Francisco State University, in developing a comprehensive set of

student learning outcomes in mathematical reasoning (Mills, et al., *Learning Outcomes for Mathematical Reasoning for the Baccalaureate Degree*, Learning Outcomes Project Final Report, California State University, January 15, 1999). That study identified nineteen different student learning outcomes deemed desirable for graduates of the CSU. (See appendix.) These nineteen learning outcomes provided a valuable starting point for this study. Recognizing that nineteen outcomes were too many to target initially, input was solicited from current GE mathematics faculty at CSU, Chico to help rank these outcomes by relevance and suitability for assessment.

Ultimately, a total of six student learning outcomes were identified as most significant and pertinent to this assessment project (Table 1, with highlighted goals corresponding to the indicated outcome). The first three outcomes are suitable for assessment by indirect means, i.e., by way of a student response survey. The last three are suitable for direct and embedded assessment. (See the appendix of *Learning Outcomes for Mathematical Reasoning for the Baccalaureate Degree* for a full description of the significance of each learning outcome.)

These student learning outcomes were particularly appropriate to this study because they are strongly correlated with goals of the GE program, as expressed in EM 99-05. These goals fall into three different categories: GE program goals, GE Core skills, and GE Mathematics (Sub-Area A4) goals. Goals relating specifically to quantitative reasoning, either directly or indirectly, are summarized in Table 2.

Table 1: Selected Student Learning Outcomes

Student Learning Outcomes in Quantitative Reasoning	GE Program Goals			GE Core Skills		GE Math Goals	
Baccalaureate graduates of CSU, Chico will be able to:	1	2	3	1	2	1	2
Outcomes 1 - 3 are suitable for indirect assessment, e.g., a student response survey.							
1. view mathematics with heightened interest, increased confidence, and less anxiety as a result of their educational experiences.		■					
2. regard mathematics as a way to think, reason and conceptualize, not simply as a set of techniques.		■			■		
3. understand and appreciate the connections between mathematics and a variety of quantitative and non-quantitative disciplines.		■	■		■		
Outcomes 4 - 6 are suitable for direct assessment, e.g., student performance on a problem.							
4. develop and apply measurement techniques to data collection, and evaluate potential sources of error, including variability and bias.	■			■	■	■	■
5. interpret, make appropriate judgments, and draw reasonable conclusions based on numerical, graphical and symbolic information.	■			■	■	■	■
6. critically evaluate quantitative information, and identify deceptive or erroneous reasoning.	■			■	■		■

Each of the selected student learning outcomes is associated with one or more of the GE goals (Table 1). For example, Outcome #2, *Regard mathematics as a way to think, etc.*, is associated both with GE Program Goal #2 and with GE Core Skill #2. It is noteworthy that all six learning outcomes together encompass all of the GE goals, i.e., assessment of all six student learning outcomes would concurrently evaluate all mathematics-related aspects of the GE program.

Much of the prior assessment of mathematics in general education has focused on indirect measures applied to individual MATH courses. In the current study, the focus was to be on direct and embedded means of assessing student performance. That is, students in a course would be evaluated by their ability to perform a task relevant to that course, where the task would also be associated with one or more overarching student learning outcomes.

Table 2: GE Goals Relevant to Quantitative Reasoning (EM 99-05)

Relevant GE Program Goals are:

1. to improve … mathematical reasoning, analysis and problem solving, and the ability to access, evaluate, and apply information.
2. to enhance general knowledge and attitudes …
3. to provide … coherence, connectedness, and commonality within broad areas of undergraduate education.

Relevant GE Core (Area A) Skills

The principal charge to this area is to provide students opportunities to learn and demonstrate:

1. effective mathematical reasoning.
2. fundamental links between thinking … and mathematical reasoning.

GE Mathematics (Sub-Area A4) Goals

Students must demonstrate:

1. understanding of one or more mathematical fields …
2. understanding of more than computational skills; they must also demonstrate understanding of basic mathematical concepts and apply these concepts to complex real world activities.

b. Determine which SLOs will be assessed, and identify course locations, tasks, and rubrics for assessment of the SLOs.

With the list of SLOs reduced to six items, the Task Force planned to seek broad input from faculty across campus to determine campus SLO priorities and possible locations for measuring the SLOs. With the goal of identifying locations for embedded assessment of the SLOs, the Task Force initially targeted non-mathematics courses likely to involve quantitative reasoning, focusing on upper-division general education courses, but also including service courses for technical majors. However, several events caused the Task Force to modify this plan.

- Informal communication with upper-division GE instructors indicated that, with just a few exceptions, there is very little quantitative reasoning in these courses. In most of the exceptions, the instructor-identified tasks were primarily low-level arithmetic, not quantitative reasoning.

- In a few cases, upper-division GE instructor-identified tasks involving the use of formulas, but even here, the tasks seldom involved even algebraic reasoning. Students simply substituted values into a formula and evaluated the result.
- It was much easier to identify higher-level quantitative reasoning tasks in service courses for technical majors (including some GE courses) and in major courses in technical disciplines. However, the Task Force and others had serious concerns about what would be measured in such courses. Would we be measuring the impact of the GE program or understanding developed in major coursework or through some other mechanism?

Consequently, by the end of fall 2005, it was determined that the assessment would be performed primarily in lower-division MATH courses with instructors willing to embed a common task. This effort was undertaken with the assumption that this would be a first step in a multi-year project. Consequently, it would be better to start with a fairly uncomplicated approach with the expectation that subsequent studies would be more sophisticated. Early in spring 2006, the Task Force crafted a probability task and a calculus task and sought faculty volunteers to imbed these tasks in their courses. This two-pronged approach allowed the Task Force to collect data from a broad student population. With very few exceptions, most students spend a part of their mathematics coursework studying either probability or calculus; in fact, Liberal Studies is the only major with students who would not typically be enrolled in at least one course in the study. To assist with future efforts, embedded assessment was also piloted in three upper-division courses, two in GE and one outside of GE. While data from the two upper-division GE courses could not be included in the study (for reasons described below), data from the upper-division non-GE course, a course taken by all engineering majors, is included. Table 3 provides details about the participating courses.

Table 3: Participation

Course	# Sections	# Students Enrolled	# Student Participants
Probability Task Results			
MATH 101: Patterns of Mathematical Thought	7	294	196
MATH 105: Statistics	8	273	157
MATH 107: Finite Mathematics for Business	7	260	221
CIVL 302: Engineering Economy and Statistics	1	66	33
Upper-Division GE	2	82	0
Probability Task Totals	**25**	**975**	**607**
Calculus Problem Results			
MATH 120: Analytic Geometry and Calculus	**5**	**153**	**100**

c. Develop instruments to be used to gather student performance data, student demographic information, and student attitudinal data.

The probability task selected is a minor modification of a task used in LEAP-M in 2002; similar problems occur naturally in a variety of Area A4 courses, increasing the pool of potential faculty participants. The calculus task was selected in consultation with MATH 120 faculty; a problem of this sort is common in all sections of first-semester calculus and such a problem frequently appears on common final exams in calculus. Each task relates to only two of the identified three student learning outcomes (5 and 6), but these two in combination with the student response survey described in the next paragraph are associated with all applicable GE goals. See Appendix C for the problems, the problem solutions and the calculus scoring rubric.

The conversation with GE mathematics faculty indicated the high priority they place on the attitudinal outcomes 1 through 3. Therefore, with the assistance of Institutional Research at CSU, Chico, the Task Force developed and administered a survey to the same students participating in the mathematics assessment tasks. (See Appendix C for the survey.) The objectives of the survey were to learn something about the math-related experiences of Chico students, to attempt to gauge their attitudes and perceptions regarding mathematics, and to collect student demographics. Moreover, the embedded task and survey were designed so that the survey findings could later be correlated, as desired, with individual student performance on the mathematics tasks.

2. Implementation of Assessment

a. The Probability Task and Survey

In a series of two multiple choice questions, students are first asked to determine the most likely outcome of five flips of a fair coin and to then explain why they chose their first answer. In early March, faculty in spring 2006 sections of MATH 101 (Patterns of mathematical Thought), MATH 105 (Statistics), and MATH 107 (Finite Mathematics for Business) were given the probability task, asked to include the probability task on an exam, and asked to administer the survey to their students. The collection of possible participants was expanded to include two upper-division GE classes and CIVL 302 (Engineering Economy and Statistics). Faculty who agreed to participate were also given a recording sheet so that they could record student responses as they graded their exams. By asking for student identification information on both the faculty recording sheets and the survey, we were able to link student performance on the task with demographic and attitudinal information about the student.

There were 975 students enrolled in the participating courses. A total of 607 students completed both the probability task and the survey; data analysis is based on this sample of 607 students. In a few cases, entire sections of student data were not used because results of either the task or the survey were not available. In one case, an instructor of an upper-division GE course modified the probability task, resulting in a task that was substantially different from that used in other courses; we were not able to include that instructor's data in our analysis. See Table 3.

b. The Calculus Task and Survey

The calculus problem is a standard related rates problem; similar problems can be found in any calculus textbook and frequently appear on final exams. Many related rates problems require some geometric reasoning before calculus can be applied. The problem selected requires only minimal geometry, an application of the Pythagorean Theorem. Since this is an open-response computational problem, the task Force worked with calculus instructors to develop a 4-point scoring rubric. The demographic/attitudinal survey was also administered to the calculus students.

All five spring 2006 sections if MATH 120 participated, resulting in three faculty participants and a potential for 153 student participants. Every class set of work was independently graded by the instructor and one of the Task Force members. After meeting to discuss two of the class

sets of scores, the Task Force determined that the rubric was clear enough to assure consistent scores across readers.

We received both problem scores and survey results for 100 students. The problem chosen was not a good fit for the CIVL 302 course. That instructor collected student performance information for a different calculus problem; those results are not used in the data analysis. See Table 3.

3. Analysis

a. The Probability Task

Table 4 displays data from the probability task and the student survey of participants in the probability portion of the study. Students in the "right" category gave both the correct answer and the correct reason for choosing that answer.

Table 4: Probability Task Data

	M101		M105		M107		CIVL 302	
right	124	63.27%	128	81.53%	154	69.68%	28	84.85%
wrong	72	36.73%	29	18.47%	67	30.32%	5	15.15%
totals	196	100.00%	157	100.00%	221	100.00%	33	100.00%
	Chi-square = 17.514, p-value = 0.0005539							

	male		female		dev math		no dev math	
right	188	67.14%	246	75.23%	164	67.21%	270	74.38%
wrong	92	32.86%	81	24.77%	80	32.79%	93	25.62%
totals	280	100.00%	327	100.00%	244	100.00%	363	100.00%
Chi-square = 4.84, p-value = 0.0278069					**Chi-square = 3.68, p-value = 0.0551349**			

	freshmen		not freshmen		math frequently use in major		math never used in major	
right	266	73.08%	168	69.14%	115	74.68%	90	67.16%
wrong	98	26.92%	75	30.86%	39	25.32%	44	32.84%
totals	364	100.00%	243	100.00%	154	100.00%	134	100.00%
Chi-square = 1.11, p-value = 0.2918646					**Chi-square = 1.62 , p-value = 0.2031**			

	confident		not confident		attitude +		attitude -	
right	362	71.83%	18	72.00%	247	75.08%	101	64.74%
wrong	142	28.17%	7	28.00%	82	24.92%	55	35.26%
totals	504	100.00%	25	100.00%	329	100.00%	156	100.00%
Chi-square = 0, p-value = 1					**Chi-square = 5.574 , p-value = 0.018229**			

The data indicate some statistically significant differences in populations:

- Females were more likely than males to have the right answer for the right reason.
- Students who reported having taken a developmental mathematics course did not score as well as those who did not require developmental mathematics.

- Students who reported a positive attitude toward mathematics (determined by responses to the prompt "I have a good attitude about math.") scored better than those with a negative attitude.
- Students in MATH 105 scored better than either MATH 101 students ($p = 0.0001613$, chi-square test) or MATH 107 students ($p = 0.009126$, chi-square test).
- Students in CIVL 302 also scored better than students in MATH 101 or MATH 107, but the difference was somewhat less significant. (E.g., for CIVL 302 versus MATH 107, $p = 0.0713799$, chi-square test).

There were no significant differences in performance between freshmen and other students. Students' mathematical confidence (determined by responses to the prompt "I can use math correctly.") had no statistical relation to their performance on the task. The difference in performance between students in MATH 101 and MATH 107 was not statistically significant. While there were some differences in performance between students in different colleges, these differences were probably a reflection of the courses the students took. For example, 75.82% of the 91 students from the College of Behavioral and Social Sciences answered both questions correctly versus 64.14% of the 145 students from the College of Business; however, almost all students from the College of Business were enrolled in MATH 107, while students from the College of Behavioral and Social Sciences were more likely to be enrolled in MATH 105.

About the chi-square test: The chi-square test is used to measure the degree of confidence that the difference in characteristics of the samples can be generalized to differences in the population, and are not due to random error. The p-value is the probability if a chi-square value at least as large as the given value. The smaller the p-value, the greater is the statistical significance of the difference in characteristics of the samples.

b. The Calculus Task

Table 5 displays some of the data collected from the calculus problem and student survey. In the displayed data, there were no statistically different levels of performance in the groups examined. While males performed slightly better than females (mean 2.92 versus 2.52, respectively), the difference was not significant. Similarly, students who reported that they would be taking additional mathematics courses in the future did somewhat better than students in their terminal mathematics class (mean 2.91 versus 2.5, respectively), but again the difference was not significant. There was little difference between the freshman and non-freshmen groups. Attitude did not appear to be related to performance on the problem.

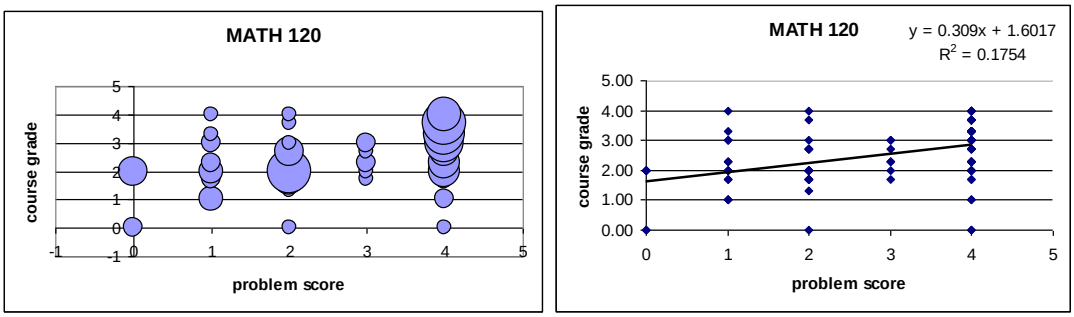
Table 5: Calculus Problem Data

score	male		female		attitude +		attitude -	
4	41	56%	9	33%	39	51%	11	48%
3	3	4%	4	15%	7	9%	0	0%
2	16	22%	7	26%	15	19%	8	35%
1	8	11%	6	22%	13	17%	1	4%
0	5	7%	1	4%	3	4%	3	13%
totals	73	100%	27	100%	77	100%	23	100%
	mean = 2.92, σ = 1.36		**mean = 2.52, σ = 1.28**		**mean = 2.86, σ = 1.32**		**mean =2.65, σ = 1.47**	
	t-test, p = 0.189				**t-test, p = 0.524**			

score	no more math		yes more math		freshmen		not freshmen	
4	10	42%	40	53%	21	43%	29	57%
3	1	4%	6	8%	5	10%	2	4%
2	6	25%	17	22%	13	27%	10	20%
1	5	21%	9	12%	7	14%	7	14%
0	2	8%	4	5%	3	6%	3	6%
totals	24	100%	76	100%	49	100%	51	100%
	mean = 2.50, σ = 1.44		**mean =2.91, σ = 1.31**		**mean = 2.69, σ = 1.33**		**mean = 2.92, σ = 1.37**	
	t-test, p = 0.197				**t-test, p = 0.401**			

There was moderate correlation between a student's score on the calculus problem and the student's grade in the class; illustrating that assessment is different from grading. The bubble graph in Figure 1 illustrates frequencies of pairs of problem scores and course grades. The second graph in Figure 1 illustrates the pairs without frequencies, but includes the regression line for the entire data set. It is important to recall that only students who completed the survey *and* took the final exam were included in the sample, resulting in a sample containing few "D" or "F" course grades.

Figure 1: Calculus Problem Score versus Course Grade



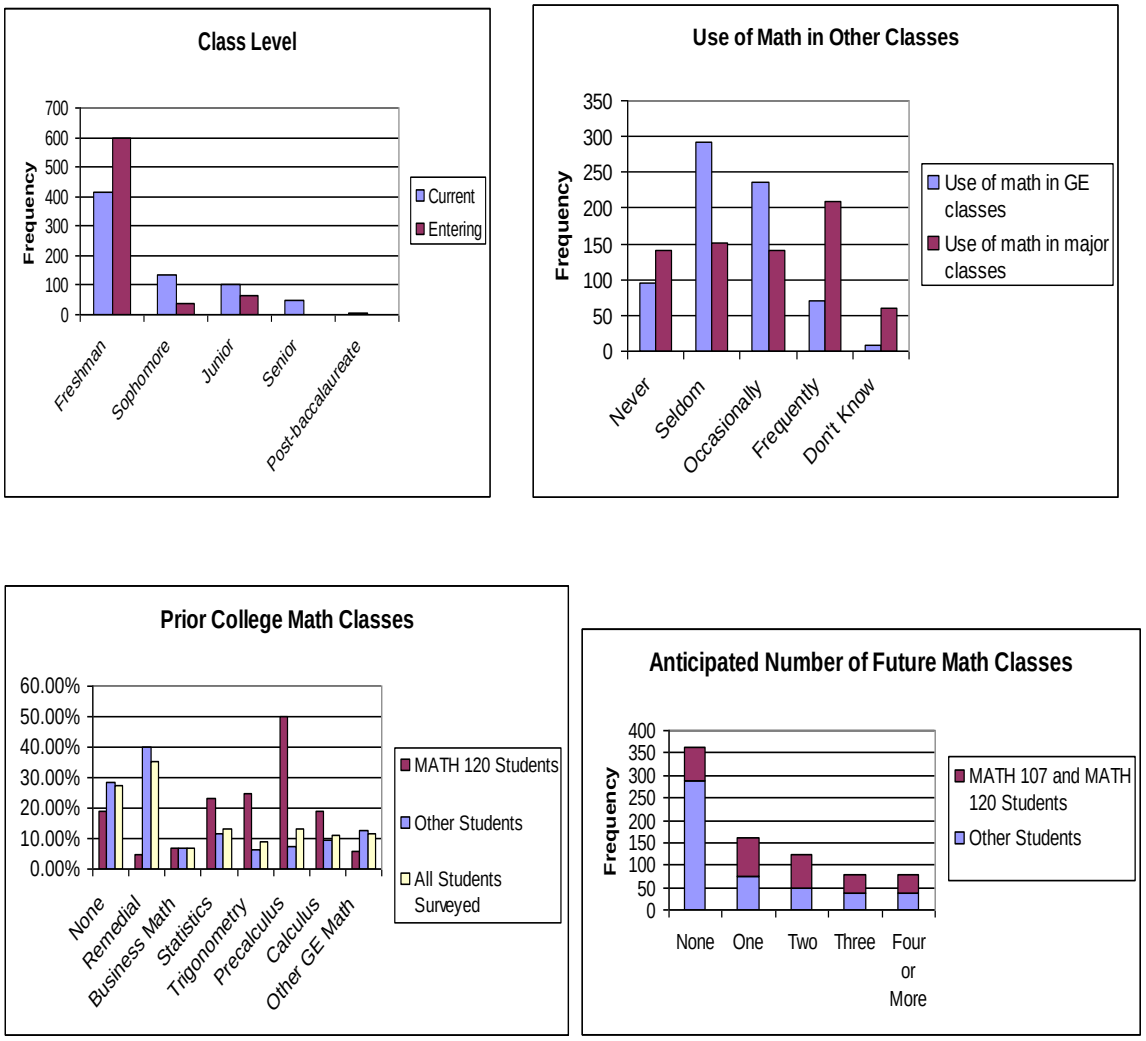
c. Survey Results

Figure 2 provides a summary of additional information collected in the student survey.

Observations:

- Since most of the students surveyed were freshmen, student perceptions about the use of mathematics in other classes are necessarily somewhat uninformed. In particular, students may think of low-level arithmetic as an example of "use of mathematics."
- Student perceptions confirm what faculty had suggested: mathematics plays at best a minor role in other general education courses.
- Excluding the MATH 120 students, most students in the sample experienced no previous college-level mathematics coursework.
- While many MATH 107 and MATH 120 students will take at least one more mathematics class, most other students are in their terminal mathematics class.

Figure 2: Survey Data



A. Conclusions and Recommendations for Strengthening QR in GE

- **Establish benchmarks for quantitative reasoning expectations.**
  This is a campus-wide task. What are the campus expectations for quantitative reasoning and at which stages should we assess student progress?

- **Embed quantitative reasoning in courses across the curriculum.**
  Are we happy with the results? If not, faculty outside of math must take on their share of the responsibility. We do students a disservice when we avoid QR simply because students find it difficult. Like writing skills, quantitative reasoning skills need to be reinforced and developed over time and in a variety of settings, both in the majors and in GE.

- **Provide and support professional development opportunities in quantitative reasoning across the disciplines, including GE.**

Develop a collection of resources for faculty wishing to incorporate significant QR into their courses. Provide workshops in which faculty can work together to develop such interdisciplinary resources, including appropriate assessment instruments.

B. Future Steps

- What is the mathematics enrollment history of a typical student? (Analyze a sample of student transcripts?)
- Begin a campus conversation about the results of this study.
- Help the campus understand that this study is only a first step is assessing quantitative reasoning and that quantitative reasoning cannot be assessed solely in mathematics courses.
- We don't know if our students leave college better able to reason quantitatively than when they entered. Do we wish to measure "value added?" How would we measure this?

C. Assessing Oral Communication in GE (Guzley, Avanzino, Johns)

Background Information

With Executive Order #595 dated November 1992 oral communication became a requirement in the CSU General Education – Breadth Requirements. At CSU Chico, Oral communications is part of the GE Core, Area A-1 (See EM 99-05.) Students may fulfill the oral communication requirement with one of two courses: CMST131 Speech Communication Fundamentals or CMST132 Small Group Communication. Both courses are housed in the Department of Communication Arts & Sciences, Communication Studies Program. The framework for instruction differs in the two classes. In the CMST131 class a professor provides one lecture per week in a large class setting. This professor is also the coordinator/supervisor for the graduate teaching assistants who conduct the balance of weekly class meetings for CMST131 in sections of approximately 25 students. Each teaching assistant is responsible for teaching from one to three sections of the class. In contrast, CMST132 is taught predominantly by experienced part-time faculty although a few sections are taught by full-time tenured or tenure/track faculty. There is a coordinator of the small group classes to ensure uniformity of instruction. In both CMST131 and CMST132 the syllabi are consistent overall as is the curriculum across sections, though there is some variation in class activities.

In addition to fulfilling Oral Communications requirements in Area A-1, EM 99-05 specifies that "[i]n every course, relevant skills of the Core must be applied as essential to the process of mastering content and making applications." This objective is further reinforced in the capstone requirement, which dictates that "[t]hemes will incorporate, build upon, and nurture skills from Area A…" As with Quantitative Reasoning, it appears that there has never been a serious attempt to ensure that Oral Communications skills are included in all GE courses, nor assess the effectiveness of these efforts. As part of the assessment process, detailed below, all instructors of GE courses in Spring 06 were contacted to elicit how Oral Communications skills were being "nurtured" in their courses. The response was disappointing and it quickly became evident that the mandate of EM 99-05 to incorporate oral communications across the GE curriculum was being honored in the breach, if at all. As with Quantitative Reasoning, a campus-wide discussion of the role of Oral communications in GE is definitely warranted.

Assessment Process

Initial stages. There were five tasks addressed by the task force fall 2005 in preparation for the actual assessment of oral communication in GE courses spring 2006:

- Review and revision of GE oral communication requirements
- Identification of spring 2006 GE courses containing an oral communication assignment
- Instructor agreement to participate in the GE oral communication assessment
- Development of oral communication rubric to assess three SLOs
- Review of rubric by participating GE instructors and Area A-1 class coordinators (CMST131 and CMST132)

There is no institutional memory of oral communication being systematically assessed at the GE level since it became a requirement in the GE curriculum. Thus, the GE Oral Communication Task Force began its task fall 2005 by reviewing EM 99-05, the most recent version of General

Education requirements at CSU Chico, with particular attention to the oral communication (Sub-Area A1) goals. The task force members found the four goals for oral communication requirements in EM 99-05 both vague and confusing in places. They revised the goals to more accurately reflect student learning goals, and also constructed student learning outcomes (SLOs) for each goal. (For a comparison of the GE oral communication goals appearing in EM 99-05 with the revised goals and newly developed SLOs, see Appendix B.) One goal and three student learning outcomes (SLOs) were selected for measurement spring 2006:

Student learning goal:

> *Students will demonstrate ability to effectively prepare for and deliver public presentations*
>
> SLO 1: *Students will effectively evaluate content for oral presentations*
> SLO 2: *Students will effectively organize content used in oral presentations*
> SLO 3: *Students will effectively deliver oral presentations*

The task force members constructed an oral communication rubric to measure these three SLOs (see appendix B). Rubrics from Mary Allen's Assessment Workshop were consulted along with rubrics from other departments on campus. The resulting oral communication rubric had three rating categories for each of the three SLOs to be assessed (organization, content, delivery). Those categories were *effective, adequate*, and *inadequate*. A complete description of each category appears on the rubric. The same rubric was used for the A-1 classes and the theme/upper division GE classes. (Note: The word "effective" in the student learning goal and the accompanying SLOs above is not intended to equate to the evaluation rating category of *effective*. Instead, the word "effective" is intended to signify that students are performing at or above a minimal acceptable level. That is, ratings of their oral communication performance for each SLO are expected to be either *effective* or *adequate*.)

Identification of Courses to be Assessed. Task force members concluded that both Area A-1 classes (CMST131 and CMST132) should be represented in the assessment and notified the coordinators of those classes they would be included in the assessment. To provide a measurement of oral communication skills beyond the freshman/sophomore level, task force members also agreed GE upper division and theme classes needed to be included in the assessment.

In December 2005 the task force sent an email to department chairs for all departments offering GE classes spring 2006 requesting confirmation of the names of instructors for these courses. Department chair responses yielded 20 classes (many having multiple sections). Subsequently, an email was sent to the instructors of these courses in late December 2005 and again in mid-January 2006 asking: 1) whether they had an oral communication assignment in the specified GE class they would be teaching spring 2006; and 2) whether they would be willing to include their GE class in the GE assessment. Ultimately, five instructors of GE upper division and theme courses agreed to participate in the oral communication assessment. One instructor withdrew from the assessment one week before the assessment began. The remaining courses represented a good mix of individual and group presentation assignments. A list of the participating classes appears in the section entitled "Assessment Data Collection" below.

All participating instructors/course coordinators were given the rubric for assessment of the three SLOs and asked whether the rubric was a good fit to their oral presentation assignment. All gave favorable comments about the rubric and agreed it would fit their assignment well.

Assessment Data Collection. Instructors/coordinators of each of the above courses identified the dates of oral presentations to be given toward the end of the spring semester and provided the task force members with the oral presentation assignment description. With these dates the task force members constructed a schedule for videotaping presentations and solicited assistants to videotape each class. Videotaping was conducted from April 27 through May 11, 2006. The instructors notified students in each of the effected courses of the purpose of the videotaping prior to commencement of videotaping. In addition, assistants responsible for the videotaping provided introductory remarks about the purpose of the videotaping upon request.

All videotaped presentations were transferred to DVD format and edited to remove irrelevant/extraneous comments occurring before or after each presentation. A few presentations were removed from the analysis for a variety of technical reasons (e.g., errors in videotaping). All videotapes were reviewed by one of two assessment team members to determine usability in the assessment. The final number of students assessed in each of the selected classes appears below.

Table 1: GE Oral Communication Assessment Spring 2006

GE Courses	Individual/Group Pres.	# Students Enrolled	# Students Assessed
Area A-1			
CMST131	Individual	410	70*
CMST132	Group	580	64*
Area E			
NFSC100H	Group (dyad)	14	14
CIVL495	Group	61	40*
Theme A			
RELS/MCGS324	Individual	47	22*
Theme O			
POLS/WMST324	Individual	38	20

*Students represented from multiple sections

Assessment Evaluator Selection. Fifteen faculty volunteered to be evaluators for the assessment team. While faculty were paired for evaluation sessions based primarily on their availability during June 2006, care was taken that no faculty member who was associated with a course included in the GE oral communication assessment was assigned to assess that class. For example, two evaluators were instructors of the classes being assessed and one faculty was the coordinator of one of the A-1 classes assessed. These three faculty were assigned to evaluate presentations in classes they did not instruct/coordinate.

Assessment Sessions. Evaluations of student presentations in the selected classes were completed in ten assessment sessions during June 2006. Evaluation sessions lasted approximately 6-7 hours and began with a norming session of approximately 45 minutes to an

hour conducted by a facilitator (member of the GE oral communication task force). The number of student presentations evaluated per session ranged from 15-34. In all but one session (Session #6) evaluators assessed presentations from only one course.

After viewing each presentation, evaluators discussed their ratings. In the event of differing ratings, the facilitator referred them to the rubric for clarification. Inter-rater reliabilities for the three oral communication characteristics assessed (organization, content, and delivery) appear in Table 2 below. Once all evaluation sessions were completed, the evaluations of student presentations were input into the STEPS system by one of the task force members to allow for analysis.

Table 2: GE Oral Communication Assessment Spring 2006
Inter-rater Reliabilities (Kappa)

Session #	Organization	Content	Delivery
1	.87	.94	.92
2	.93	.93	.84
3	.76	.79	.78
4	1.00	1.00	.83
5	.87	1.00	1.00
6	1.00 1.00	1.00 1.00	.74 .88
7	1.00	1.00	1.00
8	.92	*NR**	.92
9	.84	1.00	.92
10	.91	.71	*NR**

**NR= unacceptable reliability. Results from these sessions are included in the frequency distributions below but have been excluded from comparison tests.*

Assessment Analysis

As indicated in Table 1 above, 230 student presentations were assessed by evaluators: 112 individual presentations and 118 group presentations. Demographic information about participants appears in the table below. Frequencies were calculated for the ratings of each oral communication characteristic. Table 4 below displays those frequencies in three categories: 1) overall for presentations assessed across all courses in the assessment; 2) for presentations in the A-1 courses included in the assessment (CMST131 and CMST132); and 3) for presentations in all other GE classes included in the assessment.

Table 3: GE Oral Communication Assessment Spring 2006
Demographic Information

Class Status		Transfer Status	
Freshman	53	First Time	180
Sophomore	30	Returning	1
Junior	69	Returning Transfer	2
Senior	78	Transfer	47
Total	230	Total	230

Table 4: Frequencies for Evaluator Ratings
Three Oral Communication Assessment Characteristics

Characteristics	Overall	A-1 Classes	Other GE Courses
Organization			
Effective	200 (43%)	156 (58%)	44 (23%)
Adequate	193 (42%)	94 (35%)	99 (52%)
Unacceptable	67 (15%)	18 (7%)	49 (25%)
Total	460	268	192
Content			
Effective	129 (28%)	94 (35%)	35 (18%)
Adequate	274 (60%)	140 (52%)	134 (70%)
Unacceptable	57 (12%)	34 (13%)	23 (12%)
Total	460	268	192
Delivery			
Effective	135 (29%)	95 (35%)	40 (21%)
Adequate	255 (56%)	149 (56%)	106 (55%)
Unacceptable	70 (15%)	24 (9%)	46 (24%)
Total	460	268	192

As indicated in the table above, at the overall level, the frequencies indicate that a small number of presentations (12%-15%) fell into the *unacceptable* category for organization, content or delivery. The same was true when frequencies were calculated for the A-1 classes alone, with only 7%-13% evaluated as *unacceptable*. Approximately the same percentage of student presentations were evaluated as *unacceptable* in content for the A-1 classes (13%) and other GE classes (12%), however, this was the extent of similarities in evaluations between the A-1 classes and the other GE classes.

We would hope to see a positive shift in the frequencies for these courses across all oral communication characteristics assessed; that is, a higher level performance in the upper division and theme GE classes than in the A-1 classes. As indicated in Table 5, however, the percentage of presentations assessed as *effective* in all three assessment characteristics is noticeably less for the other GE classes assessed than for the A-1 courses. In addition, the percentage of presentations assessed as *inadequate* in the other GE courses is approximately triple the percentage in the A-1 classes for the assessment characteristics of organization and delivery.

In summary, overall in both the A-1 classes and the other GE courses assessed most presentation evaluations indicate students have at least an *adequate* oral communication skill level in the three characteristics assessed (organization, content, and delivery). Unfortunately, however, the evaluations also reflected there was little, if any, improvement in student oral communication skills from the time students took one of the two A-1 classes to the time they reached upper division/theme GE classes (predominantly juniors and seniors), and in some ways their skills appear to have decayed. To confirm whether the differences in evaluation ratings discussed above were statistically significant, the Mann-Whitney two-group comparison test was run (see Table 5 below). Across all three oral communication characteristics evaluated, there was a significant difference between the ratings for the A-1 classes and the other theme/upper division GE classes included in the assessment.

Table 5: Nonparametric Comparison
A-1 Classes and Four Theme/Upper Division GE

	Organization Score	**Content Score[a]**	**Delivery Score[a]**
Mann-Whitney U	15236.000	17404.000	15785.000
Asymp. Sig. (2 tailed)	.000	.005	.000

[a]Unreliable rating sessions identified above excluded from comparison

Given that the two A-1 classes approach oral communication in different venues (individual versus group presentations), we computed frequencies for the evaluator ratings across the three oral communication characteristics for CMST131 and CMST132 to compare evaluator ratings.

Table 6: Frequencies for Evaluator Ratings for CMST131 & CMST132
Three Oral Communication Assessment Characteristics

Characteristics	**CMST131**	**CMST132**
Organization		
Effective	65 (46%)	91 (71%)
Adequate	61 (44%)	33 (26%)
Unacceptable	14 (10%)	4 (3%)
Total	140	128
Content		
Effective	34 (24%)	60 (47%)
Adequate	76 (54%)	64 (50%)
Unacceptable	30 (22%)	4 (3%)
Total	140	128
Delivery		
Effective	44 (31%)	51 (40%)
Adequate	78 (56%)	71 (55%)
Unacceptable	18 (13%)	6 (5%)
Total	140	128

The Mann-Whitney two-group comparison test was used to determine any significant differences between the two courses on the three oral communication characteristics assessed. There were

significant differences in the evaluator ratings between the two A-1 classes across all three oral communication characteristics rated (see Table 7 below).

Table 7: Nonparametric Comparison
CMST131 and CMST132 Evaluations

	Organization Score	Content Score	Delivery Score
Mann-Whitney U	6641.000	6128.000	7801.000
Asymp. Sig. (2 tailed)	.000	.000	.039

Conclusions

- This oral communication assessment provides us with the first view of our effectiveness in meeting our student learning goal and three related SLOs.

- For the sample of students included in this assessment of oral communication, we currently appear to be successful in meeting our goal and three SLOs. The majority of presentations evaluated achieved ratings of *adequate* or *effective* across the three assessment categories.

- The A-1 classes (CMST131 and CMST132)—which serve predominantly freshmen and sophomores—appear to be doing a good job of addressing the three SLOs evaluated in this assessment, with 87% -93% (depending on the assessment characteristic) performing at or above minimal acceptable standards.

- The CMST132 class had significantly more evaluations in the *effective* category and fewer in the *inadequate* category than did the CMST131 class indicating an examination of how each of these two classes address the three SLOs is needed to bring evaluations in future assessment into closer alignment.

- It is disheartening to see in this assessment that students at the junior/senior level are performing no better (and in some cases worse) than their counterparts in the A-1 oral communication classes, particularly given the importance of oral communication skills in their future careers, community, and family life. These findings indicate we need to know more about the extent to which oral communication is stressed in GE upper division/theme classes as well as in the various majors on campus.

Future Steps

- Work to refine the oral communication rubric for the student learning goal addressed in this assessment. It worked well for this assessment but a few comments from evaluators indicated additional clarification and extension of some rating categories would be useful.

- Work with coordinators of the CMST131 and CMST132 classes to ensure that clarified GE oral communication goals and associated SLOs are being addressed consistently and similarly in the two courses with the intent to raise the CMST131 evaluation level to that of CMST132. Integrate use of rubric developed by GE Oral Communication Task Force

into these courses. Use DVDs of presentations collected for this assessment as training tools for instructors in both courses.

- Explore with department chairs in departments where GE upper division and theme classes are housed the extent to which oral communication is a requirement in these GE classes, as well as the extent to which such requirements are consistent with the clarified GE oral communication goals and associated SLOs. Strive to identify a core of oral communication skills students build across GE and major classes while acknowledging variation beyond this core is inevitable and acceptable.

- Work with chairs, administrative bodies and other appropriate constituencies to develop a plan for how oral communication should be addressed in GE theme/upper division classes. Minimally, plan should include such things as which GE classes will address oral communication requirements, in what types of assignment (e.g. individual versus group presentations) it will be addressed, and use of rubric in assessing the assignment.

- Access from AURA and department/program chairs the number of majors on campus that have an oral communication student learning goal/outcome. Explore the consistency of those goals/outcomes with the GE oral communication goal/outcomes. Examine results of oral communication SLOs assessed spring 2006 to determine if comparisons can be made to GE Oral Communication Assessment results.

- Schedule a follow-up oral communication assessment in 2-3 years to examine the extent to which recommendations have been implemented and successful.

- Meet with GE Oral Communication Task Force members and others involved on the assessment team to debrief the assessment; specifically, develop a record of procedures that worked well in the spring 2006 oral communication assessment as well as those that need to be improved, and to identify how they might be improved.