Georgia State University
General Education Core
Student Learning Outcomes Assessment Report

General Education Outcome Assessed:
Critical Thinking in Area D Natural Sciences

Combined results from
Biology
Chemistry
Anthropology and Geography
Geology
Physics and Astronomy

Prepared by
Sarah Pallas, Chair (Biology)
Barbara Baumstark (Biology)
Davon Kennedy (Chemistry)
Paul Franklin (Chemistry)
Jeremy Diem (Geography)
Seth Rose (Geology)
Brian Thoms (Physics and Astronomy)

Date: March 28, 2005

Assessment of Critical Thinking in the Natural Sciences

SUMMARY

Goals: The charge of the Area D subcommittee of the Provost's SACS assessment team was to develop and implement assessment tools for the ability of students exiting Gen Ed Area D courses to think critically. The first goal was to develop an operational definition of Critical Thinking that could apply to all courses in Area D. That definition is outlined in the *Description of outcomes* below. Second was to determine how each course had been teaching and testing critical thinking. Any mismatch in the two was then used to inform strategies for changing the course to improve outcomes. Finally we developed end-of-term testing instruments that we hoped could be applied across sections of the same or similar courses.

Results: In the case of Chemistry and Physics, nationally normed assessment instruments already exist, and this has been of great benefit to our data collection process. This semester has seen an increased efficiency of data collection, for the most part. In cases where problems were encountered, action has already been taken. Overall the Area D courses are adequately teaching and assessing Critical Thinking. The subcommittee members are to be congratulated for their efforts.

Future Plans: Challenges that remain include getting buy-in to the process on the part of department chairs and instructional personnel. There is a need for chairs to communicate the importance of this effort to those involved in it, and to find a way to reward productive participation. In addition, communication between different instructors for the same course, and use of the same assessment tools, would greatly increase the reliability of the data. In addition, we have outlined GSU's need as well as a national need for peer-reviewed assessment instruments in the fields of Biology, Geology, and Geography. An NSF-funded opportunity to develop such an instrument for Biology has become available and will be piloted in Bio 2108 in the fall.

1. **Description of outcomes: a student that completes Area D will be able to**
    a. formulate appropriate questions and testable hypotheses for research;
    b. effectively collect appropriate (empirical) evidence;
    c. apply and integrate principles and concepts to analyze problems within specific core areas;
    d. appropriately evaluate and interpret claims, arguments, evidence and hypotheses;
    e. use the results of analysis to appropriately construct new arguments or alternate hypotheses and formulate new questions.

2. **Description of ways courses contribute to goal**
    Courses in the natural sciences employ critical thinking in many ways, including the development of analytical thought processes through working of new problems using learned formulae or facts, the development of skeptical analysis of scientific results reported in the lay and scientific literature, synthetic thinking

using learned information to predict what has not yet been learned, asking scientific questions and formulating alternate hypotheses, designing experiments to test hypotheses, carrying out those experiments, and interpreting data.

3. **Description of assessment methods** (these are possible assessment tools; appended are descriptions of specific assessment tools used in each department, and the tools themselves have been collated should the Gen Ed committee wish to see them.)
    - **a.** Entrance interviews/surveys
    - **b.** Selected exam and/or quiz questions
    - **c.** Writing assignments
    - **d.** Lab notebooks
    - **e.** Research projects
    - **f.** Formal assessment tools, e.g. standardized exams
    - **g.** Exit interviews/surveys

    Our goal is a 75% achievement score, that is, at least 75% of the students should perform at the criterion level that has been set by the department.

4. **Description of data collection and analysis procedures**
    We performed a full-scale implementation of our assessment procedures in Fall term 2004. Each academic unit provided a summary of their methods and outcomes. Where possible, performance outcomes are expressed as % achievement of criterion level. Results will be used to inform improvements to assessments in the Spring term.

5. **Description of how results will be used to suggest improvements**
    Subcommittee members in each department examine the results, looking at outcome in relation to goals, and, for courses taught by more than one instructor, comparing across courses. The data were not identified according to instructor. This helps to avoid the negative connotations associated with evaluation of instruction by individual faculty and focuses the effort where it belongs, on curriculum design across all instructors in each department. If improvement is needed, additional critical thinking exercises will be incorporated in lectures, labs, and/or projects, along with additional training of TAs.

6. **Goals for the future**
    The main difficulty is instructor compliance with the goals of the Area D team. The lower level courses in the different departments are taught not only by tenure track faculty, but also by part-time temporary faculty that may have little investment in the process. Some chairs and subcommittee members have been successful in encouraging these faculty to get on board, and some have had less success. We suggest that the institution find a more effective way to encourage investment in the process. Keys to success are clarifying goals, reducing the workload for faculty assessment efforts, and allowing flexibility in assessment tools. Although all 5 disciplines fall under the category of Natural Sciences, the

disciplines are quite different in the way they are taught and in the tools that are nationally available, and it is important to consider this in planning efforts.

TABLE 1: Courses assessed

DEPT	**Course #**					
Biology	Biol 1107K	Biol 1108K	Biol 2107K			
Chemistry	Chem 1152K	Chem 1211K	Chem 1212K			
Geography	Geog 1112K	Geog 1113K				
Geology	Geol 1121K	Geol 1122K				
Physics	Astr 1010K	Astr 1020K	Phys 1111K	Phys 1112K	Phys 2211K	Phys 2212K

Assessment of Critical Thinking in Biology

Courses Assessed: Biol 1107K, 1108K, 2107K
Assessment materials for 2108K were never received. Although it would be ideal to have data for each of the 4 introductory biology courses, it is not essential to evaluating the success of our efforts in the department.

Assessment Tools:
The faculty involved in the fall term were asked to
•employ how and why questions during lectures and labs,
•give information and ask students to use it to make a knowledge-based prediction,
•give students an observation and ask them to develop a testable hypothesis about it,
•ask students to put "2 and 2 together" and combine information in new ways,
•ask students about what sort of evidence was necessary for a particular scientific discovery to be accepted,
•ask students to draw conclusions from a graph or set of data

The possible assessment tools that could be used were:
•Entrance survey to assess prior knowledge of concepts and scientific method and of CT skills.
•Lab exercises that involved CT
•Quizzes with embedded CT questions
•Exams with embedded CT questions
•Term papers that depended on CT ability.
•Exit survey to assess developed knowledge of concepts and scientific method and of CT skills.

Summary of Results:
Biol 1107K. The 1107K students were assessed in the laboratory, with the head instructor formulating the assessment questions and overseeing the grading. Two questions were used in the assessment, which was presented as part of the lab practical exam.

The results were: 1107K MWF section:#1: 79%. #2: 84%
1107K TTh section: #1: 80%. #2: 81%
Total # of students assessed: 212

Biol 1108K. The instructor summarized the results for the 1108K critical thinking assessment. He and the co-instructor used two essay questions:

Biol 2107K. The two faculty administered two questions at the end of the semester.

Overall Conclusions:
1. There is a wide divergence in what is considered to be “critical thinking.” The sub-subcommittee’s bias is that a critical thinking question should be formulated to

include as much of the content knowledge as possible that is needed for the answer. Many of the questions used for this assessment could not have been answered without prior content knowledge, so they would seem to be measuring content knowledge as well as, or possibly instead of, critical thinking.

2. There is also a wide divergence in what might be considered to be a satisfactory response to the questions. Since the "'pass rate" is based on the percentage of satisfactory responses, it would be important that all participating instructors use similar evaluation parameters.

Changes Planned for Spring 2005: Of the available tools, only the questions embedded in the final exam were reported. Two of the faculty did not do any assessment at all. These were the 2108K faculty as noted above. Thus a major future goal will be to get equivalent participation of all faculty, and to develop a set of questions that can be used in common across different sections of the same course. All introductory biology instructors (majors' and non-majors') should administer the same (or equivalent) sets of questions, and should use the same rubric for assessment. Since so many of the instructors in these classes are our lecturers, and since they have regular meetings scheduled, this might be a task that they want to become involved in. Ways to encourage this are being explored with the department chair and lead instructional faculty.

Assessment of Critical Thinking in Chemistry

Courses Assessed: Chem 1151K, 1211K, 1212K

Assessment Tools: The American Chemical Society provides national-level exit exams for all of the area D courses within the chemistry program. A representative faculty committee for area D was formed and functioned by forming subcommittees. Subcommittee recommendations provided a set of critical thinking questions from each of these exit exams. The departmental committee was composed of tenured faculty and lecturers that teach the particular courses as well as ad hoc members (instructors) that teach in area D. The departmental committee voted unanimously for both the set of questions and the expected number of correct answers for each course. After the exit exams were given, we tallied how many students met the expected standard of 2 of 7 or better correct on the selected critical thinking questions. The percentage of students meeting this standard is also given. The expected outcome was based on the Department of Chemistry Learning Outcome rubric submitted to the Provost's Office prior to the fall semester. The following data was collected.

Summary of Results:

Chemistry 1151 (first semester Survey of Chemistry)

	Raw Score
Expected Performance	2 of 7 on 1151 exit exam
Actual Class Performance	74 of 75 scored 2 or better (98.7 % were successful)
Average Score	2.8 of 7

Chemistry 1211 (first semester General Chemistry for Majors)

	Raw Score
Expected Performance	2 of 8 on 1211 exit exam
Actual Class Performance	119 of 125 scored 2 or better (95.2 % were successful)
Average Score	4 of 8

Chemistry 1212 (second semester General Chemistry for Majors)

	Raw Score
Expected Performance	2 of 8 on 1212 exit exam
Actual Class Performance	117 of 120 scored 2 or better (97.5 % were successful)
Average Score	3.3 of 8

Overall Conclusions: The expected score for each area D course is 2 correct out of the questions selected and all students should be able to meet this standard. We found that our students do meet the expected standard in every course.

Changes Planned for Spring 2005: We will continue to monitor the performance of our students in these courses but no changed to our methods are planned at this time.

G. Davon Kennedy
Associate Professor

Paul Franklin
Director of Undergraduate Studies

Assessment of Critical Thinking in Geography

Courses Assessed: Geog 1112, 1113

Assessment Tools: An assessment of critical thinking by students in Geography 1112 (Introduction to Weather and Climate) and Geography 1113 (Introduction to Landforms) has been made through evaluations of exam responses and group essays. The assessment was intended to involve two lecture courses of both Geography 1112 and Geography 1113; however, insufficient participation by three of the four instructors resulted in only a single Geography 1112 course with complete and valid data. The other Geography 1112 course had biased data, while one of the Geography 1113 courses had partial – yet problematic – data. The remaining Geography 1113 course had completely insufficient data. Consequently, the following assessment is focused on the aforementioned Geography 1112 course, with supplemental assessment involving the Geography 1113 course with partial data.

Geography 1112 Assessment Methods

Ten exam questions were given to students throughout the Geography 1112 course. Each question was designed to assess student performance on at least one of five components of critical thinking: 1) discussion of central concepts and relationships; 2) presentation of evidence for accepting an argument; 3) presentation of evidence for doubting an argument; 4) explanation for the acceptance of a conclusion; and 5) identification of additional information needed to make a sound judgment. Student responses were assessed with the aid of a rubric with possible values being 1 (poor), 2 (fair), 3 (good), and 4 (excellent).

Group essays were constructed by student groups in response to a summary of atmospheric research on precipitation enhancement that appeared in a peer-reviewed journal. Students were provided with several forms of data and then asked to address the following questions/statement in their essays:

- What are the central concepts and relationships of the research topic?
- What evidence exists for supporting the conclusion of the researchers?
- What evidence exists for doubting the conclusion of the researchers?
- Explain why you accept or reject the conclusions of the researchers. If you neither accept nor reject the conclusion, then explain why you have that stance.

Each question/statement represented a component of critical thinking, and responses from each group were assigned scores than ranged from poor (1) to excellent (4). Since expressing thoughts through writing is a central component of critical thinking, this exercise might be a better critical-thinking assessment tool than are the exam questions.

Geography 1113 Assessment Methods

Ten exam questions were given to students throughout the Geography 1113 course. Similar to the Geography 1112 questions, these questions were designed to assess the major components of critical thinking. Unfortunately, student responses to only four questions (1, 2, 4, and 5) are available.

The format of the group-essay assignment was modeled after the Geography 1112 assignment (see Appendix); however, instead of being provided with data and actual

research results, the students were given a hypothetical situation. In their essays, students were asked to address the aforementioned four questions/statements. Possible scores ranged from poor (1) to excellent (4).

Summary of Results:

Geography 1112 Results

Based on the student responses to the exam questions, the over-all critical-thinking ability of undergraduate students in Geography 1112 can be rated as fair (i.e. composite score was a 1.7) (Table 1). The exam scores also revealed that a large proportion of the students possess poor critical-thinking ability. The percentages of students in each of the four critical-thinking performance categories are as follows: poor (32%), fair (64%), good (4%), and excellent (0%). As a whole, nearly all students were poor to fair at critical thinking, while few to none of the students were good to excellent critical thinkers.

Results from the group essays echoed the results from the exam-question responses: Geography 1112 students are fair critical thinkers (Table 2). The composite critical-thinking score was a 1.9. The congruence between results from the exam questions and the essays provides enhances validity to the overall results. The percentages of students/groups in each of the four critical-thinking performance categories are as follows: poor (15%), fair (67%), good (15%), and excellent (3%). These results combined with the exam-question percentages indicate the following: nearly one-quarter of the students are poor critical thinkers; and only one to two students in a class of 100 students are excellent critical thinkers.

Close examination of the group essays revealed some interesting findings concerning student thinking. Some of those findings are as follows:

- more than one-quarter of the groups definitely did not understand the concept of wind direction, which is a critical piece of foundation knowledge provided to the students in the textbook and in lectures and laboratory exercises;
- nearly one-quarter of the groups had substantial contradictions in their papers;
- more than one-quarter of the groups accepted the researchers' findings yet provided stronger arguments for rejecting rather than accepting the findings. A common response from those groups was that information from satellites – which were the data upon which the researchers made their conclusions – is accurate, even though the accuracy of the satellite data was not provided to the students; and
- only three of the eleven groups that neither accepted nor rejected the researchers' conclusions explicitly stated that additional information was needed before a decision could be made.

On a positive note, five of the thirteen groups that rejected the researchers' conclusions presented a well-supported hypothesis of where precipitation enhancement should have occurred. In addition, several groups did note that the study period – which was three years – was not long enough.

Geography 1113 Results

Student responses to the exam questions indicated that the overall critical-thinking ability of undergraduate students in Geography 1113 is fair. The results are somewhat contaminated by the nature of the exam environment: students were given total access to their textbook and notes while taking the exam. These testing parameters were not conducive to student thinking, for students assumed that all the answers were readily available in their notes or textbook or both. Therefore, the exam scores should be treated as a loose cross-validation of the Geography 1112 exam scores. One notable result was the poor performance of every student on Question 4. None of the students recognized the plate-tectonics aspect of the question. Nearly all the students listed forms of evidence of continental drift.

The writing assignment provided more reliable information on the critical-thinking ability of Geography 1113 students than did the exam-question responses, and the over-all critical-thinking rating for the students was fair (i.e. composite score was a 1.9) (Table 4). It should be noted that the assignment was slightly easier than the Geography 1112 assignment. For example, Geography 1113 students did not have to interpret any data. The percentages of students/groups in each of the four critical-thinking performance categories are as follows: poor (10%), fair (80%), good (10%), and excellent (0%). Similar to the findings for Geography 1112, most of the students are poor to fair critical thinkers.

Conclusions: This critical-thinking-assessment exercise has revealed that most students in Geography 1112 and Geography 1113 possess fair critical-thinking abilities. Possibly more important, the assessment revealed that it is not unreasonable to conclude that up to one-quarter of the students in those courses are poor critical thinkers. The distribution of critical-thinking ability is positively skewed: the lack of good and excellent critical thinkers is alarming.

Changes Planned for Spring 2005: The assessment tools described in this report were successful, but they were by no means devoid of problems. Major shortcomings of the assessment procedure included the following: 1) insufficient participation by instructors; 2) small number of exam questions; and 3) assessing individual performance based on group work. An ideal situation would have the following characteristics:

- assessments conducted only in courses taught by tenure-track faculty members;
- the inclusion of more writing-centered critical-thinking questions on exams throughout the semester; and
- the request of individual essays as opposed to group essays.

Nevertheless, additional resources (e.g., dedicated teaching assistants) would be needed to implement most of those changes. Future assessments should be comprised of at least ten questions and group essays in each lecture session of Geography 1112 and Geography 1113.

Assessment of Critical Thinking in Geology

Courses Assessed: Geol 1121, 1122

Assessment Tools: Most of the four faculty (all tenure-track) used a battery of exam questions to assess Critical Thinking. One used a single, more open-ended essay question. Different instructors used different exam questions, thus results cannot be compared directly across sections.

Results: Student performance on the critical thinking assessment questions varied greatly (i.e. between 40-90%). In general, students performed slightly poorer on the critical thinking questions than on the remainder of these examinations (these comparative data, however, have not been included in this report). The wide range of variation not only existed between the different professors who took part in this initiative but was very noticeable within single examinations given by the same professor. In general, the Geology Department was not overly dissatisfied with student performance; however, there is certainly room for improvement.

Summary: The Department of Geology has now participated in Area D of the University's Critical Thinking Initiative for two semesters. Each faculty member in the department who teaches GEOL 1121 and GEOL 1122 (Introductory Geology I and II) is responsible for assigning a critical thinking component to their assessments. Most of the faculty assign critical thinking questions as part of their class examinations that are used to determine the student's grade in the course. There are no routine pre-test/post-test analyses in that limited classroom time and heavy faculty workloads do not readily allow for such procedures. The critical thinking questions are typically designed to see how well students can take a concept that has been taught, and apply it in a way that was not specifically covered during the lecture. Most of the questions represent small but meaningful steps in this regard. The Geology faculty believe that our entire curriculum in Introductory Geology has been designed in such a manner that students must critically assess how our planet works. Students also use critical thinking skills in the Geology laboratories that are separate from the lecture portion of the class. However, we have not included questions from laboratories as part of our assessment for the committee.

Changes Planned for Spring 2005: At this point we believe the format for assessing critical thinking that we are currently using in our core courses (GEOL 1121 and GEOL 1122) will not undergo radical changes. Every semester a different set of professors from the Geology Department will be assigned to teach these introductory courses. In that the department relies upon the individual professor to teach and grade a course, there is no standard format for evaluating "critical thinking" in these classes. Some professors give special assignments while others include critical thinking-type of questions as a routine portion of their classroom examinations. There is also a critical thinking component to our labs. These questions were not included in this summary (Fall, 2004); however, we will try to include some of our laboratory work on the assessment for Spring 2005. A smaller class size will allow for some pre-test and post-test, essay type of evaluations to be included in next semester's evaluations.

Assessment of Critical Thinking in Physics and Astronomy

Courses assessed:

Fall 2004:

Astr 1010 (3 lectures, 13 sections total) Astronomy of the Solar System
Astr 1020 (3 lectures, 13 sections total) Stellar and Galactic Astronomy
Phys 1111K (4 lectures, 12 sections total) Introduction to Physics I
Phys 1112K (2 lectures, 6 sections total) Introduction to Physics II
Phys 2211K (4 lectures, 12 sections total) Principles of Physics I
Phys 2212K (2 lectures, 6 sections total) Principles of Physics II

Assessment Tools:

During the summer session, an in-depth study of a representative sample of students was conducted. Critical thinking questions were identified, graded, and the correlation with overall course grade was examined. It was clear that the questions which require critical thinking were more challenging than other test questions, and that the critical thinking abilities of the students were correlated with overall success in the course.

For the fall semester, we chose to broaden the assessment, with the goal of including all students in most sections of the courses. All those teaching these sections were asked to participate, and twelve faculty members reported meaningful results, representing about 800 students.

Each of these faculty members included two targeted critical thinking questions on their final exam. The questions are included below. For most students, they cannot be answered directly from memorized material, but require some thought.

One goal of our approach is to standardize the assessment tools used in each class so that we can compare results between instructors and from term to term. By using only final exam questions we can use the same question over a number of semesters. In addition to standardizing the questions, we also established grading rubrics to be used by each instructor.

Assessment was also performed in the laboratory for the physics courses. Quizzes were given containing identified critical thinking questions to lab students during the next to last week of the labs. Students were given extra credit for this quiz. Since the chief difference in the 1000 and 2000 level classes is not the physics content but the mathematics applied to it, and since questions can be written that do not use advanced mathematics, students in Phys 1111 and Phys 2211 took the same quiz as did students in Phys 1112 and Phys 2212. A random sample of papers was used for the critical thinking analysis.

Results:

Astr1010 and Astr1020

Percent Correct

Astr 1010	Section 1	Section 2	Section 3
Question 1	26%	22%	33%
Question 2	34%	36%	36%
Astr 1020	Section 4	Section 5	Section 6
Question 3	39%	35%	23%
Question 4	40%	18%	37%

Note: A small group of non-science honors students scored 44% and 77% on the two Astr 1010 questions.

Phys1111K

Average Score

	Lecture 1	Lecture 2	Lecture 3	Lecture 4
Question 1&2	23%	25%	20%	37%

Phys1112K

Average Score

	Instructor 1	Instructor 2
Question 1	70%	29%
Question 2	16%	26%

Phys2211K

Average Score

	Instructor 1	Instructor 2
Question 1	22%	34%
Question 2	24%	55%

Phys2212K

Average Score

	Instructor 1	Instructor 2
Question 1	81%	
Question 1b		41%
Question 2	17%	39%

Physics Labs

Average Score

	Phys1111	Phys2211	Phys1112	Phys2212
Question 1	58%	42%	81%	64%
Question 2	48%	61%	58%	25%
Question 3	58%	56%	21%	28%
Question 4	60%	56%	42%	52%
Question 5	52%	51%	33%	24%

Overall Conclusions:
It is clear that the questions requiring critical thinking are more challenging than other test questions. This is not surprising since they require both an understanding of the science and also critical thinking and reasoning skills. Since the goal is to assess critical thinking in the context of a specific core area, the average score on these critical thinking questions may always be somewhat below the average score for all questions.

Since we have implemented a assessment scheme that will continue to use the same questions and the same grading system, a comparison from term to term should give meaningful data on changes in the critical thinking skills of our students. These changes can then be correlated to changes in our curriculum or teaching methods.

Changes Planned for Spring 2005:
Comparing the results of the same question from instructor to instructor reveals large variations in the results. Although some of this variation is undoubtedly due to differences in the instruction and the differences in the class make-up, some of it is most likely due to variations in grading. This is not surprising since Fall 2004 was the first semester that we used standardized questions. To make comparison across instructors and semesters more meaningful, we will work to make the grading of the questions more uniform.