

General Education Outcomes Scoring Session Report

Written Communication and Critical Thinking Skills

May 24, 2004

Prepared by Joan Hawthorne

On Tuesday, May 24, a group of 19 UND faculty gathered to review papers written by graduating seniors and score them on two key general education skills: written communication and critical thinking. This report summarizes the results of that scoring session.

Student Papers

All papers used for rubric development and/or scoring were solicited during the past year from faculty who were teaching courses that they or their colleagues considered to be, formally or informally, capstone sorts of experiences. This is in keeping with the AAC&U recommendation that our students' "best work" should be evaluated when determining outcomes of general education programs. Papers for this assessment were submitted from students in ten departments (Anthropology, Communication, Electrical Engineering, History, Honors, Management, Nursing, Political Science, Recreation and Leisure Services, and Teaching and Learning).

In the case of some departments (i.e., Communication, Teaching and Learning), papers were submitted as part of a final portfolio process built into the program and required for graduation. Papers included in these final portfolios were widely disparate and it was not possible to be entirely consistent in the kinds of papers taken from these portfolios (e.g., a case study written in a T&L class might have served as a sample from one portfolio while a research paper from a content area class was a sample from another). In other departments, papers were written by students in capstone courses (e.g., History, Anthropology) but were not part of a portfolio process. Some departments contributing papers were able to gather papers from courses taken by senior students but not necessarily described as "capstone" courses (e.g., Management, Nursing, Recreation).

One difficulty in this process was caused by the variability in papers submitted. Although faculty in all cases were told that the purpose of the process was to look at written communication and critical thinking skills of graduating seniors, typical assignments vary considerably from one part of campus to another. Furthermore, papers submitted ranged from less than a page in length up to 20 or even 50 page papers submitted as culminating projects in writing-intensive disciplines. This tremendous variability in type of paper undoubtedly contributed to problems with scorer reliability (described below).

Rubric Development

Prior to the scoring session itself, rubrics were developed and tested by a small team of faculty representing a broad disciplinary range (Joan Hawthorne, Education/Writing Program; Kim Fink, Visual Arts; Dave Pierce, Chemistry; Lori Robison, English; Paul Sum, Political Science and Public Administration, and Anne Walker, Teaching and Learning). Sample rubrics currently in

use at several other institutions were reviewed as part of the rubric development process, and the rubrics developed by the group, although tied to UND's gen ed goals and faculty perceptions of the UND academic environment, were quite similar to those used at other institutions for similar purposes. Final versions of the two rubrics are attached to this report.

Scoring Procedures

Upon the recommendation of the rubric committee, faculty divided into two teams, one to score for written communication (led by Joan Hawthorne) and the other to score for critical thinking (led by Anne Walker). Team members began by reviewing the rubric appropriate for their task, and then each team conducted a norming process. To norm, faculty were given an assortment of sample papers, and then were asked to score a first set of two papers. Faculty shared their first sets of scores and discussed their rationale, in terms of the rubric, for assigning a particular score. Where possible, faculty tried to move toward agreement on a single score, so that a given paper could serve as a marker for a certain level of proficiency. Additional papers were scored during the norming, using similar procedures.

With a possible range in total score of 0 to 9 on the written communication rubric, a variation of up to 1.5 was considered acceptable; with a range from 0 to 6 on the critical thinking rubric, a range of up to 1.0 was considered acceptable. A third reader was to be used any time scores assigned by the first two readers were outside those tolerance limits. It became clear during norming that reaching agreement on a score, even within these ranges, would be quite difficult. In the written communication group, faculty agreed to move from norming to individual scoring despite this problem, recognizing that inter-rater reliability was likely to be lower than desired, but also maintaining the integrity of individual scorers' ratings. In the critical thinking group, faculty dealt with norming discrepancies by pairing themselves so that a highly consistent scorer (as determined by scores in the group norming process) was paired with a less consistent scorer, with the two individuals reading and scoring as a team. Members of a critical thinking scoring team consulted as they each assigned scores to an individual paper, and a third reader was employed only if, after discussion, their individual scores were still outside of the pre-determined range of acceptability.

After the first few papers were scored collaboratively, both groups moved into individual paper reading and scoring, using the two different procedures described above. Each paper was scored by two readers, independently or as a team, with a third reader used in those cases where the first two scores were outside predetermined tolerance limits. All paper scores reported below are the mean of scores assigned by the two or three readers of a given paper.

Written Communication

The following faculty scored 53 student papers for written communication: Jeanne Anderegg (Honors), Virgil Benoit (Languages), Joan Hawthorne (Education and Coordinator of the University Writing Program), Dorothy Keyser (Music), Martha Potvin (Biology and VPAA/Provost), Libby Rankin (English and Director of the Office of Instructional Development), Peggy Shaeffer (Teaching and Learning), Kathy Sukalski (Biochemistry and Molecular Biology), and Paul Sum (Political Science and Public Administration). Of the 53 papers scored, 5 scores

were assigned collaboratively as part of the norming process, 21 represent the mean of three readers' scores, and 27 represent the mean of two readers' scores.

The mean score on the student papers was 5.21 (on a scale of 0 to 9), with a range of scores from 0.5 to 9.0. The actual distribution of scores is as follows:

0.0 to 0.4	0 papers
0.5 to 1.4	3 papers
1.5 to 2.4	3 papers
2.5 to 3.4	6 papers
3.5 to 4.4	8 papers
4.5 to 5.4	9 papers
5.5 to 6.4	10 papers
6.5 to 7.4	2 papers
7.5 to 8.4	2 papers
8.5 to 9.0	10 papers

However, it may be more useful to consider scores in broader groupings, given the following considerations: (a) our purpose is to examine the degree to which graduating students seem to be achieving proficiency in select general education skills by the time of graduation, so the rubrics were not developed with an eye to making fine distinctions in precise level of accomplishment (i.e., the intention was not to develop an instrument that would finely distinguish between levels of achievement in the same way as grades are expected to within the context of a class); (b) we did not achieve a high level of inter-rater reliability among individual scorers, which would have given more credibility to fine scoring distinctions; (c) differences in types of assignments, papers, and disciplines, especially combined with differences among scorer backgrounds, reduce the precision possible in scores and mean that the suggestion of a high level of precision would be misleading. Therefore, summary scores were regrouped into broader categories that can be expected to provide more credible, and therefore more useful, information regarding the degree of skill in written communication demonstrated, in these particular papers, by UND students.

As a rough division, it seems reasonable that student scores below 3.0 represent a generally disappointing level of achievement, since such a low score means that the paper received a score of less than 1.0 on at least one criterion, and earned a mean score of less than 1.0 on the three criteria. Individual criterion scores in the 0 - 0.5 range clearly indicate unsatisfactory demonstration of skill on the criterion in question. In contrast, total scores above 6.0 demonstrate a highly satisfactory level of skill, since a student must achieve a 2.5 or 3.0 score on at least one criterion, combined with a mean of above 2.0, in order to earn a total score above 6.0. Distinguishing between satisfactory and unsatisfactory scores in the middle portion of the scale, where most of our students understandably fell, is considerably more difficult. One way to consider the level of skill demonstrated in these mid-range papers is simply to divide the remaining scores down the middle and consider whether students tended to be above or below the mid-line (4.5). Using this method of division, about half of the remaining students fall above the mid-line and half fall below, with five students falling exactly on the line, as can be seen by the groupings below:

0.0 to 2.9	highly unsatisfactory	7 papers
3.0 to 4.4	low mid-range	13 papers
4.5	precisely mid-range	5 papers

4.6 to 6.0	high mid-range	12 papers
6.1 to 9.0	highly satisfactory	16 papers

Thus it appears that a slim majority of graduating seniors (28 out of 53 students, or 53%) produced work that was generally regarded by a diverse group of faculty readers as satisfactory or highly satisfactory in terms of written communication skills. A small minority (7 students or 13%) produced highly unsatisfactory work, and a total of 20 students (38%) produced work that might be described as slightly to highly unsatisfactory. Or, looked at another way, 7 students (13%) produced highly unsatisfactory work and 16 (30%) produced highly satisfactory work – but 30 students (57%) fell somewhere in the middle. As a graph, these numbers would yield a clear positive skew but with a still significant number of graduating seniors producing work in the highly unsatisfactory category.

In addition to total paper scores, scorers reported individual scores for each criterion. Papers scored collaboratively during the norming portion of the session were not given individual criterion scores, which means that 119 individual criterion scores were reported. The distribution of times each score was used by a scorer is as follows:

Score	Criterion #1 (sense of purpose)	Criterion #2 (guidance for readers)	Criterion #3 (clarity and conventions)
0.0	12 uses	11 uses	14 uses
0.5	2 uses	1 use	0 uses
1.0	32 uses	36 uses	28 uses
1.5	7 uses	9 uses	8 uses
2.0	29 uses	28 uses	40 uses
2.5	5 uses	6 uses	7 uses
3.0	32 uses	28 uses	22 uses

Two points of interest can be seen from a cursory review of the data. First, written communication scorers very heavily used whole numbers for scoring, resorting to a half-point option only in a small percentage of the cases. Second, scores on the third criterion (clarity and conventions) tended to run slightly lower than scores on the first two criteria, although not so much lower as to disproportionately skew total scores.

Critical Thinking

The following faculty scored student work using the critical thinking rubric: Matt Cavalli (Mechanical Engineering), Anne Kelsch (History), Ed Kolodka (Chemical Engineering), Mike Mann (Chemical Engineering), Doug Munski (Geography), Mike Nitz (Communication), Lana Rakow (Communication), Fred Remer (Atmospheric Sciences), Anne Walker (Teaching and Learning), and Margaret Zidon (Teaching and Learning). Scorers completed work on 76 papers; 6 were scored collaboratively as part of the norming process, 61 were scored by a two-person scoring team, and 9 received three scorings.

The mean score for papers scored using the critical thinking rubric (scale of 0 to 6) was 3.45, with a range of 1.0 to 5.5. The distribution is as follows:

0.0 to 0.4	0 papers
0.5 to 1.4	3 papers
1.5 to 2.4	14 papers
2.5 to 3.4	17 papers
3.5 to 4.4	19 papers
4.5 to 5.4	18 papers
5.5 to 6.0	5 papers

As with written communication scores, it may be fairer and, therefore, more useful to break scores down into broader categories.

0.0 to 1.4	highly unsatisfactory	3 papers
1.5 to 2.9	low mid-range	20 papers
3.0 to 4.4	high mid-range	30 papers
4.5 to 6.0	highly satisfactory	23 papers

Using this distribution as the basis for analysis, it becomes clear that only 3 students (4% of the papers scored) demonstrated a highly unsatisfactory level of achievement. In contrast, 23 students (30%) demonstrated a highly satisfactory level of achievement. Of the 65% of mid-range papers, 26% (20 papers) could be described as low mid-range, and 39% (23 papers) as high mid-range. If graphed, this distribution would show a very strong skew to the positive side, although it's important to note that 4% of students demonstrated a highly unsatisfactory level of achievement and 30% demonstrated achievement that could be described as at least slightly unsatisfactory (low mid-range or below).

Scores for individual criteria were also reported. Excluding papers scored collaboratively during the norming (which were not given individual criterion scores), there were 147 individual criterion scores, which were distributed as follows:

Score	Criterion 1 (purpose)	Criterion #2 (analysis)	Criterion #3 (resolution)
0.0	4 uses	1 use	7 uses
0.5	22 uses	22 uses	36 uses
1.0	38 uses	60 uses	50 uses
1.5	49 uses	44 uses	35 uses
2.0	34 uses	20 uses	19 uses

In comparison to the criterion scores for written communication, it is immediately apparent that scorers used the half-points with great frequency, presumably because they were working with a narrower scale (0, 1, 2 rather than 0, 1, 2, 3) and felt the need for half-points to more finely distinguish levels of achievement. It is also apparent that scores on the first criterion are higher than those on the other criteria, and scores on the second criterion are higher than scores on the third. In other words, students seemed to achieve the highest levels of performance on the purpose criterion, and their poorest performance generally was on the resolution criterion, with performance on analysis somewhere in the middle.

Conclusions

This is the first time any cross campus scoring of this sort has been conducted at UND, which means that it is difficult to draw firm conclusions. However, this pilot scoring session still yielded several findings that should be useful to us in considering the success of our general education program.

Findings regarding the scoring process:

1. Common definitions of “critical thinking” and “written communication,” while necessarily broad, are shared among many faculty. Members of a rubric committee, despite widely varied backgrounds, were able to reach consensus regarding criteria that could be considered as defining traits demonstrating the necessary skills. Members of the scoring group, with faculty of even more widely varying backgrounds, were mostly satisfied with the criteria they were asked to use and proposed few significant changes in the rubrics even after attempting to use the rubrics to assess achievement on a wide range of sample papers.
2. Although there appears to be general agreement about the nature of these general education skills, there is less agreement on how to apply a rubric describing the various levels of achievement. This created problems in inter-rater reliability.
3. With a project of this sort, inter-rater reliability can be expected to be difficult to achieve given the wide range of kinds of papers scored, the wide range of scorer backgrounds, and the necessity of working within a very limited time frame. Early in the planning process, the idea of conducting a statistically-based scoring was rejected due to insurmountable difficulties in obtaining truly random samples of student work plus the extensive amount of time that would be demanded to score a statistically meaningful number of papers. In a scoring process that was not expected to yield statistical generalizability, problems with rater reliability are not insurmountable; should there ever be a desire to conduct a large-scale generalizable study of student achievement of these goals, reliability concerns would need to be addressed.
4. If this process is to continue into future years, there would be great value in encouraging faculty conversation and faculty development around criteria described on the rubrics (or revised versions of rubrics, if significant revisions should be desired). In addition, students should be made aware that the university expects to see achievement of certain levels of skill on these criteria by the time students graduate from UND.
5. If the process is to continue, thought should be given to regularizing collection of student work that can serve as evidence of achievement of these and other gen ed goals.
6. Faculty participated in the scoring session because of their interest in and commitment to student learning. These kinds of conversations about student learning and outcome are important for us as teachers, and should be encouraged in whatever ways are feasible. This could include financial incentives, “counting” toward service, enhancing the faculty development aspect of the session, etc.

Findings regarding student achievement:

1. Most of the graduating seniors whose work was included in this study were able to demonstrate a satisfactory or highly satisfactory level of achievement in the designated skill (53% on written communication and 69% on critical thinking). On critical thinking especially, this indicates an overall high level of achievement.
2. Despite the widespread success with which students were able to demonstrate these general education skills, significant numbers failed to demonstrate adequate skill achievement. Of these, a small percentage (13% for written communication and 4% for critical thinking) demonstrated highly unsatisfactory skill levels, and many more (totaling 38% for written communication and 30% for critical thinking) demonstrated somewhat unsatisfactory levels of achievement.
3. The vast majority of these students demonstrated skill levels that fall very much in the middle of the achievement range. That is, 57% of students demonstrated mid-range achievement in written communication, and 65% of students demonstrated a mid-range level of achievement on critical thinking.
4. Scores for student performance in critical thinking were generally satisfactory; scores for student performance in written communication were somewhat disappointing. Additional data regarding students' written communication skills should be collected, either from other sources or from a second scoring session conducted in another year. If it is determined that student skills in this area are weak, there may be a need for faculty to renew their efforts to teach (and demand) higher levels of writing proficiency on written work produced across the curriculum – and for more emphasis on written communication within general education coursework as a foundation for those higher demands.
5. On the critical thinking rubric, students achieved higher scores for performance on purpose than on the other criterion. Faculty might be encouraged to help students learn how to more effectively analyze and resolve issues in their writing.
6. On the written communication rubric, skill levels demonstrated on the three criteria were nearly equal, although there were fewer students who excelled on the criterion of clarity and conventions. Faculty across the disciplines might be encouraged to hold students to uniformly high standards of writing.

University of North Dakota
General Education Scoring Guide for Effective Communication

	3	2	1	0
<i>Sense of Purpose</i>	Writer is sophisticated in his/her ability to signal purpose to reader. Focused and incisive, the paper reflects a writer with a strong sense of what s/he is trying to do or say. The various sections of the paper make sense together and the writer has indicated the larger implications or importance of the written work.	There is a controlling idea that holds the paper together. While the paper might not contain a traditional “thesis statement,” there is a strong sense that the writer has a clear vision of his/her project. The various parts of the paper fit with the writer’s sense of project. The writer has a reason for writing.	While there may be a sense of purpose that holds the paper together, it is often very broad. This lack of focus may result in a very general project; the paper may therefore rely more on summary than on analysis. Writers in this category may discover a sense of purpose as they write, but they haven’t revised the entire paper to reflect this new focus.	Paper seems disjointed or incoherent. Relationship between different sections is unclear or relationship comes only from “stream of consciousness” or tangential connections between ideas. Writer may seem to be engaged in many different projects at once.
<i>Guidance for Readers</i>	The writer demonstrates a sophisticated awareness of his/her reader. Writing flows smoothly from one idea to another. The writer has taken pains to assist the reader in following the logic of the ideas expressed and has taken pains to explain and develop his/her ideas.	Generally, readers feel that the writer has helped them to understand his/her project. Sequencing of ideas within paragraphs and connections between paragraphs make the writer’s points easy to follow. The writer has found a way of developing his/her ideas, providing the reader with the examples, illustrations, and explanations necessary to understand the project.	At times, readers may feel lost and unable to follow the writer’s train of thought. Writer needs to improve sequencing of ideas within paragraphs and needs to do more to explain the connections between paragraphs. Paper may include examples and illustrations but often lacks explanations of the relevance of those examples; or paper may include explanations without the examples or illustrations the reader needs to fully understand.	The lack of connections between ideas makes reading and understanding difficult. The lack of examples, illustrations, and explanation makes understanding difficult.
<i>Clarity and Conventions</i>	Clarity of ideas is enhanced by writerly expression. Writer seems to be in command of conventions of writing and uses them to rhetorical advantage. Visual presentation of written work, formatting and/or documentation is polished.	Word and sentence choices convey meaning clearly. Writer generally controls conventions of writing. Visual presentation of written work, formatting and/or documentation is consistent and generally follows conventions. Occasional missteps in use of conventions or in presentation do not impede understanding.	Word choice and/or sentence structure gets in the way of clear communication. Writer’s inconsistent use of conventions of writing is distracting to the reader and interrupts comprehension. Visual presentation of written work, formatting and/or documentation is inconsistent and interrupts understanding.	Reader must occasionally guess at reader’s meaning. Writer’s control of conventions of writing is uncertain enough to impede comprehension. Visual presentation of written work, formatting and/or documentation is inappropriate and impedes understanding.

University of North Dakota
General Education Scoring Guide for Critical Thinking

	2	1	0
<i>Purpose</i>	<ul style="list-style-type: none"> • Provides a clearly articulated statement that defines the main question, problem or issue • Demonstrates the relevance or importance of the question, problem or issue • Selects and attributes appropriate sources of literature, evidence or academic dialogues in terms of amount and balance 	<ul style="list-style-type: none"> • Identifies several pertinent questions, problems or issues, but does not established a focused position or direction • Identifies only a marginal context for the main question, problem or issue • Selects sources, evidence or dialogues that only support the main conclusions 	<ul style="list-style-type: none"> • Shows no clear sense of purpose. No question, problem or issue is concisely stated or even implied. • Demonstrates no clear sense of importance or relevance of the main question, problem or issue • Omits important sources of literature, evidence or dialogues, whether these sources support or challenge the main conclusions
<i>Analysis</i>	<ul style="list-style-type: none"> • Uses reliable and accurate information/ evidence that is relevant to the main question, problem or issue addressed • Disassembles and reassembles relevant information in an accurate, critically-oriented, deep way producing a synthesis of the material • Demonstrates a logical progression of thought throughout the writing reflecting information and ideas that are well-structured and prioritized 	<ul style="list-style-type: none"> • Information/ evidence is of marginal relevance to the main question, problem or issue addressed in the writing • Full synthesis is not achieved. Information evidence is not disassembled and reassembled in a fully accurate, critical or deep way • Progression of ideas does not fully unfold in a logical manner, or the ideas presented are not well-structured in relation to one another 	<ul style="list-style-type: none"> • Information/ evidence does not clearly relate to the main question, problem or issue addressed in the writing • Synthesis of information is not attempted or lacks accuracy, critical assessment or depth. Attempt to disassemble and reassemble information/ evidence is absent or executed poorly • Logical progression of ideas is absent and no attempt to prioritize information is taken
<i>Resolution</i>	<ul style="list-style-type: none"> • The artifact contains a clearly articulated argument that is strongly resolved with supporting information and conclusion/s. • Conclusions provide a thorough and relevant summary of the question/issue/problem and its analysis. • The artifact poses realistic and insightful solutions and/or broader implications. 	<ul style="list-style-type: none"> • The artifact contains an argument that is minimally resolved with supporting information and conclusion/s. • Conclusions provides a brief or incomplete summary of the question/issue/problem and its analysis. • The artifact poses solutions and/or broader implications that may be simplistic or slightly unrealistic. 	<ul style="list-style-type: none"> • Artifact does not make an argument or does not resolve it with supporting information and adequate conclusions. • Conclusions are missing or irrelevant/inappropriate to the question/problem being explored. Conclusions may not be based on analysis and supporting information. • The artifact does not pose solutions or broader implications, or the solutions/implications may be extremely inappropriate.