The University of North Dakota
Office of Essential Studies

Report

CAPSTONE ASSESSMENT WORKSHOP:
ASSESSMENT OF SENIOR STUDENTS’ WORK IN CRITICAL THINKING AND WRITTEN COMMUNICATION

Conducted December 12, 2011

Tom Steen, Director of Essential Studies
Joan Hawthorne, Director of Assessment & Regional Accreditation
Anne Kelsch, Director of Instructional Development
Under the joint auspices of the Assessment Director and the Essential Studies Office, we held the first-ever Essential Studies Capstone Assessment workshop in December, 2012. Our hope is to make this an annual event, to capture senior student learning data as they complete their work in their ES Capstone courses. Not only will this allow us to see what’s going on in terms of student learning at the culmination of our students’ general education, but it will also provide us with data that we can use for the overall program assessment of ES.

We decided to do this first try with volunteer faculty. Our logic in doing so was that, in testing out the procedures, it might be best to use faculty who have a strong interest in ES. Therefore, we simply asked faculty to help out; we drew from three kinds of faculty: instructors of ES “C” courses, faculty who had been involved in developing or participating in the ES program, and members of the ES and Assessment Committees.

We also tapped volunteers for samples of student work. We asked instructors of recent and current “C” courses—Spring & Fall, 2011—to give us samples of their students’ projects or papers, if they were willing to do so. (And if they were willing and able to get their students’ permission for us to use their work (anonymously and confidentially).

We selected two of the four ES goals for assessment: a) Thinking and Reasoning, and b) Communication. We chose these because they are most frequently used goals in the validation of the current ES Capstones. Furthermore, because most “C” courses identify these two goals, we focused the assessment work this time on Critical Thinking (a subset of T&R) and Written Communication. (Our plan is to rotate the assessment focus among the four goals, and their sub-goals, over a multi-year assessment cycle, probably 3 years.)

To manage the event, the Essential Studies Office (Tom Steen and Sandy Brown) handled the communication—invitations, confirmations, solicitation and receipt of student work, logistics. Joan Hawthorne, with Tom, handled the sampling strategy, organized the procedures, and supervised the data collection. Joining Joan as facilitator of the event was Anne Kelsch; they each facilitated one of the two groups of faculty, who were divided by the two ES goals selected for assessment.

This report is designed to share the results of this first ES Capstone event. We provide some detail on the procedures and methods used, to provide a template for future such events. We also provide both quantitative and descriptive results, along with a brief commentary on what the results imply about our students’ learning across campus on these two ES goal.

~Tom Steen, Joan Hawthorne, and Anne Kelsch, with Sandy Brown
METHOD

The Event

The first Capstone Assessment Workshop was held on the Monday of exam week at the end of the fall semester of 2011: December 12. We invited people—primarily faculty—to serve as reviewers, and we reviewed a large number of “student work products” (papers, project reports, etc.) that were provided by instructors of Essential Studies (ES) Capstone courses.

The reviewers were divided into two teams—one team for each of the two learning goals under this year’s review: Critical Thinking (a sub-goal under the Thinking & Reasoning goal) and Written Communication (a sub-goal under the Communication goal). Each team was facilitated by a veteran ES leader with experience in assessment of student work products.

The event started at 11:00am, included lunch, and concluded at 3:30pm (total time: 4.5 hours). We started with a review of the two rubrics and a norming session, using a few of the student papers selected for norming practice. After the facilitators determined that the teams were reliable using the rubrics, they moved on the main process. Papers were numbered for identification—see also sampling method, below—and then passed so that each paper was reviewed by more than one reviewer. In most cases, each paper was reviewed by at least two different reviewers (a third reviewer was used if the first two were not close in agreement). Scores were kept on sticky notes, with the papers and the scores turned in to the facilitator who kept a master record of the scores for each team. With about half-hour left, the facilitators engaged their teams (independently) in a general discussion about the quality of the student work products and the particular learning goal that they worked with.

Following the event, the three of us—the two facilitators and the director of ES—collected the results and saved them for later analysis.

The Teams of Reviewers

Besides the two facilitators, the two ES goal teams numbered 21 persons total.

Critical Thinking team: 9

Written Communication team: 12

Reviewers represented different kinds of involvement in ES: instructors of “C” courses (8), ES Committee members (8), Assessment Committee members (2), instructors of other ES courses (3), and staff from the Office of Instructional Development (2).
All reviewers volunteered their time for this first-time event. In the future, we plan to address ways to make it possible for other faculty to take part and share the load of program assessment work.

**Sampling Method: Selection of Student Work**

We employed a sampling strategy was “purposive.” M.Q. Patton uses the term to indicate selecting a sample according to a rational strategy that gains a representative cross-section of the population ([Qualitative Research and Evaluation Methods, 3/e. Sage, 2002](#)), which in this case was student papers, or work products. We ended up reviewing about 50 papers in each team, and we conducted about 2.5 reviews on each paper (we aimed to complete two reviews on each paper, i.e., two different reviewers reading each paper; we completed a third review in those cases where the first two reviewers recorded total scores beyond our pre-determined range of acceptability).

<table>
<thead>
<tr>
<th>Student Work Products</th>
<th>Number Scored at Workshop</th>
</tr>
</thead>
<tbody>
<tr>
<td>Critical Thinking</td>
<td>54</td>
</tr>
<tr>
<td>Written Communication</td>
<td>41</td>
</tr>
</tbody>
</table>

**Norming Method: Calibrating Reviewers’ Judgment**

A small number of typical papers--representing the variety of disciplines, types of assignments, and degrees of excellence--were selected for use in calibrating. Copies of the papers for norming were distributed to all scorers and time was provided for independent reading and preliminary scoring of the first two. After all scorers had read and scored those papers, scores were recorded and group members discussed the reasons for their scores, in relation to the relevant rubric.

The objective was to reach consensus on appropriate scores for the scored papers and to identify papers which could serve as “markers” for what a score of, e.g., 3 on a given criterion, really meant. Once group members were able to reach agreement, another paper was scored and discussed. After three or four papers, as scorers began to internalize the scoring system and reach a shared understanding of what each score meant, the calibrating was complete and the actual individual scoring commenced.

**GOALS AND RUBRICS**

As we mentioned in the overview, for this particular assessment event, we decided to focus on two of the Essential Studies goals: a) Thinking and Reasoning, and b) Communication. We chose these because they are most frequently used goals in the
validation of the current ES Capstones. Furthermore, because most “C” courses narrow
down these two goals, we focused the assessment work this time on Critical Thinking (a
subset of T&R) and Written Communication. Both of these sub-goals have already-
developed rubrics, which are posted on the ES website (see Faculty section: Rubrics for ES
Assessment). These rubrics—developed locally at UND, using an organic, emergent,
inductive process—represent the University’s commitment to a tradition of qualitative
assessment going back over two decades. Not only were they already in place and
familiar to a number of faculty, but they also demonstrate UND’s desire to examine
student work directly (instead of more indirect, second-hand kinds of
reviews/assessments).

All the UND rubrics can be found at: http://und.edu/academics/essential-
studies/rubric.cfm

Assessment Plan. Our plan is to rotate the assessment focus among the four goals,
and their sub-goals, over a multi-year assessment cycle, probably three years:

1. Critical Thinking and Written Communication
2. Information Literacy and Diversity
3. Quantitative Reasoning and Oral Communication

We plan to use the ES capstone courses as one of the main data collection points since
they are at the end-point of students’ undergraduate experience. However, in some
cases, it will make more sense to draw student work from other courses (e.g., Special
Emphasis courses for a particular learning outcome). Thus, since there are few capstones
that have been approved for the goal of diversity, we will draw student work products
from the special emphasis courses in Global (“G”) and U.S. (“U”) Diversity.

RESULTS: CRITICAL THINKING

One of the two faculty teams reviewed the student papers with respect to their
expression and use of critical thinking, using the rubric designed to assess that particular
learning outcome. In this section, we present those results, first in quantitative terms,
then in descriptive terms.

Results for Critical Thinking

[Insert CT Results Table about here]
Analysis and Discussion: Critical Thinking

There were 128 scorings of 57 work products completed by the critical thinking scorers. The rubric for critical thinking includes three criteria, and each criterion was scorable within a range of 0-4. A perfect score on all three criteria would yield a total of 12.

Scores were analyzed in three different ways. First, individual paper scores were totaled and a mean across the scorings of that paper was calculated. For example, if a scorer rated a paper as meriting a 2.0, 1.5, and 2.0 on the three criteria, the total would be 5.5. If the total for a second scorer were 6.0, the mean score would be 5.75. Mean total scores for critical thinking were calculated for each paper scored. The majority of the work products (25 out of 57 scored) yielded total score means in the 6.1-8.0 range, which is just above the mid-point of the 12 points possible. The score distribution was as follows:

<table>
<thead>
<tr>
<th>Range of Scores (from CT Rubric)</th>
<th>Number of Papers Scored</th>
</tr>
</thead>
<tbody>
<tr>
<td>0-2.0 (lowest)</td>
<td>1</td>
</tr>
<tr>
<td>2.1-4.0</td>
<td>7</td>
</tr>
<tr>
<td>4.1-6.0</td>
<td>13</td>
</tr>
<tr>
<td>6.1-8.0</td>
<td>25</td>
</tr>
<tr>
<td>8.1-10.0</td>
<td>9</td>
</tr>
<tr>
<td>10.1-12.0 (highest)</td>
<td>2</td>
</tr>
</tbody>
</table>

This distribution shows that most papers were judged as partially meeting desired standards for critical thinking. However, considering that the papers scored are work products prepared by senior level students, it is somewhat disappointing that only 2 of the 57 scored in the very highest range (10.1-12). More concerning is the number in the lowest two ranges and the fact that a total of 21 out of 57 papers were below the mid-range (6.0) mark.

A second level of analysis looks at the scores on individual criteria. Criterion scores demonstrate that, as might be expected, students were more successful at demonstrating Knowledge & Comprehension than they were at Analysis & Synthesis, and Evaluations & Conclusions seemed to be the most challenging elements of critical thinking. No scores in the 0-0.5 range were given for Knowledge & Comprehension, while there were 4 Analysis & Synthesis scores in that range and 19 Evaluations & Conclusions scores. Conversely, 34 scores on Knowledge & Comprehension were in the top range (3.5-4.0),
13 on Analysis & Synthesis, and only 1 on Evaluation & Conclusions. If the range is divided into thirds, there are 9 Knowledge & Comprehension scores in the bottom third compared with 32 Analysis & Synthesis scores and 54 Evaluation & Conclusions scores. The top third includes 87 Knowledge & Comprehension scores, 44 Analysis & Synthesis scores, and 21 Evaluation & Conclusions scores.

Individual criterion scores can also be considered in terms of the median possible score (2.0). There were 89 scores above the median on Knowledge & Comprehension, and only 10 were below (29 scores of exactly 2.0). On Analysis & Synthesis, there were 46 scores above the median and 34 below (48 at the median), and on Evaluations & Conclusions, there were 24 above-median scores, 59 below-median, and 37 at the median.

Both ways of analyzing individual criterion scores show that student scoring on critical thinking is significantly stronger on the more basic dimensions of thinking and weaker on the most advanced dimension. However, both ways of analyzing scores also support the conclusion that senior students are developing and can demonstrate some critical thinking skills, even though they may fall short of our aims for the more complex levels of thought.

A third level of analysis looks at the total number of criterion scorings in each of three areas of the range. Criterion scores from 0-1 represent scores at the lower third of the range, 1.5-2.5 represents the mid-range, and 3.5-4.0 represents the top third. A total of 95 criterion scores fell within the lower third, 129 were mid-range, and 152 were top-range.

Each way of looking at the scores provides a different perspective on student strengths and weaknesses, but two conclusions seem clear. First, the majority of senior students are successfully demonstrating some kinds of critical thinking skills, with only 9 scores assigned, e.g., in the bottom third of the range on Knowledge & Comprehensions – and even those nine scores were at a 1.0 level rather than a 0 or 0.5.

Second, there are also significant numbers of senior students who, near the point of graduation, fail to demonstrate a level of critical thinking likely to be deemed acceptable according to the scorers. Of the 57 papers scored, mean scores for 8 were in the lower third of possible scores and only 11 were in the top third. On Evaluation & Conclusions, the score of 0 was assigned by at least one scorer 19 different times.

These results demonstrate that there is significant room for improvement despite the achievements demonstrated.
RESULTS: WRITTEN COMMUNICATION

Similar to the procedure used the assessment of the critical thinking papers, one of the two faculty teams reviewed the student papers with respect to their use of written communication. As in the other group, this team used the UND-developed rubric designed to assess WC. In this section, we present those results.

Results for Written Communication

There were 90 scorings of 42 papers for written communication, but scores for a computer science paper were removed from the data because it was viewed as inappropriate for scoring on some of the criteria, thus yielding a skewed (partial) total score. The written communication rubric contains four criteria and each criterion could be assigned a score in a range of 0-3. Therefore, as with critical thinking, a perfect score totaled 12.

The same methods of analysis were used for written communication as for critical thinking. The first level involved looking at mean total scores, considering all scorings completed for the paper, for each of 42 papers. The distribution of total scores was as follows:

<table>
<thead>
<tr>
<th>Range of Scores (from WC Rubric)</th>
<th>Number of Papers Scored</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 – 2.0 (lowest)</td>
<td>2</td>
</tr>
<tr>
<td>2.1 – 4.0</td>
<td>1</td>
</tr>
<tr>
<td>4.1 – 6.0</td>
<td>13</td>
</tr>
<tr>
<td>6.1 – 8.0</td>
<td>20</td>
</tr>
<tr>
<td>8.1 – 10.0</td>
<td>4</td>
</tr>
<tr>
<td>10.1 – 12.0 (highest)</td>
<td>2</td>
</tr>
</tbody>
</table>

The distribution shows that most papers were viewed as partially meeting the four criteria. Only three of the mean scores were at 4.0 or below (bottom third of the range), which is good news. However, one of those papers was at a mean score (after two entirely consistent scorings) of 0.0, clearly indicating a highly unsatisfactory level of achievement for a senior. Although 26 of the 42 papers (more than half) were in the upper half of the range, there were 16 in the lower half with 4 of those 16 exactly at 6.0 (the precise mid-point for total mean score).
Analysis of individual criterion scores provides an additional perspective. Scorings reveal that each criterion was quite well demonstrated by some students. Perfect scores of 3.0 were assigned to Sense of Purpose 9 times; Synthesis, Analysis, & Evaluation 12 times; Guidance for Readers; 7 times, and Clarity & Conventions 13 times. Conversely, 0 scores were assigned on Sense of Purpose 7 times; Synthesis, Analysis, & Evaluation 7 times; Guidance for Readers 4 times; and Clarity & Conventions 5 times.

Breaking criterion scores down by median (below 1.5, at 1.5, and above 1.5) provides a different perspective on criterion scores. Sense of Purpose was scored below median 24 times, at median 11 times, and above median 44 times. Synthesis, Analysis, & Evaluation was scored below median 33 times, at median 8 times, and above median 47 times. Guidance for readers was scored below median 37 times, at median 7 times, and above median 46 times. Finally, Clarity & Conventions was scored below median 34 times, at median 5 times, and above median 49 times.

The pattern demonstrated through both means of analysis is that scorings demonstrate some degree of success (considerably more scores at and above the median than below). However, the pattern also shows significant numbers of students scoring at the very bottom end of each of the four criteria, which is somewhat concerning for graduating seniors.

A third way to examine the data is by looking at the distribution of scores across the range of possibilities. With seven possible scores that could be assigned by a scorer on any criterion, it is not possible to break scores down into thirds or quarters. However, it is possible to consider the total number of scores assigned which are below median (below 1.5), at median (1.5), and above median (above 1.5). The total number of below median scores was 128, 31 median scores were assigned, and 196 were above median.

Overall, this analysis of written communication scores shows that there is significant room for improvement in this area and the need for improvement extends across the four criterion areas. The fact that there is no single area of urgent need may in some ways create special challenges for improving written communication competencies. The data suggest that individual students may have patterns of strengths and weaknesses that are quite varied.

However, the results also show that very few students are extremely weak across the board, with only three students who might be considered as falling into that category (i.e., total mean score of 4.0 or below). This number constitutes less than 1% (actually about 0.7%) of the total number of students whose work was scored. So while there is considerable room for improvement and efforts will likely need to span all four criteria, there is also some reason for satisfaction with written communication scores.
COMMENTS FROM THE REVIEW TEAMS

Notes from the Critical Thinking Team

Anne Kelsch, facilitator

Following the reviews of papers, Anne asked the group for general comments and reactions to the work that they had examined.

With respect to the student work and learning that was exhibited in these papers:

1. When a student product was strong in “Evaluation & Conclusions,” that appeared to flow from strengths in the other two categories of the rubric (i.e., Knowledge & Comprehension, Analysis & Synthesis).
2. Writing skill seemed to correlate with stronger CT. That is, when the writing was strong, the CT was also strong; when the writing was weak, so was the CT rating.
3. One common error that the team noted: the student provided a lot of information but did not appear to critically think through its value toward the paper’s main theme or key concepts. Bits of information were not well connected to each other or the main idea. (Several members of team echoed this.)
4. Another main theme of the team’s discussion was that there was a glaring difference among the set of the papers in the quality of Critical Thinking exhibited.
5. That led to a key question: is this difference something that we see across the campus in CT, beyond this set of papers? Is this quality difference something that is also reflected in all the ES Capstones (not simply in the sample of courses that submitted work for the workshop)?
6. That question, in turn, led to a discussion of: How do we work to improve the quality of CT with our own students, at any level?

With respect to the assessment review process itself:

The team suggested that the best papers to review for the purpose of program review are works that come from “major” assignments.

At the same time, shorter papers are easier to review in the relatively short time frame that the workshop allowed. Longer papers (e.g., senior theses, lengthy design project reports) are good to examine and in keeping with the purpose, but practically, they are so time-consuming to review that only a few student work products of that length would be able to be reviewed (in this day-long, retreat-type format). Using shorter papers will allow more reviews of different student work products, thus enlarging the sample to be reviewed.
Although the work was hard and the day—first day of exam week—was not the best, the team felt that reviewing the student work was a very good experience. One benefit cited was getting to see the level of work done in other departments, other majors different than their own. A second benefit was listening to colleagues’ judgment of strong/weak CT.

Notes from the Written Communication Team

Joan Hawthorne, facilitator

There was a big range in terms of student performance in writing. Many papers were uneven in sections, e.g., some sections were reasonably well-written and others were not. In fact, organization of sections was also weak on many of the papers. Synthesis of ideas was also more challenging for student writers of the papers reviewed.

As is typical for this kind of review, readers noted that there is a lot of variability in assignments and that variability may account for some of the less satisfactory work that was seen.

Readers noted that it may be packing too much into one course to combine the Advanced Communication course and the capstone. Perhaps the A course should be taken prior to the capstone so that culminating projects would have the benefit of the A learning as well as the semester of the capstone for reinforcement of disciplinary writing principles.

Finally, questions were raised about whether students really see the importance of and need for good writing. Some scorers noted that student self-reporting indicates a high level of awareness of that need, but there was skepticism based on perceptions that there were many papers which appeared not to demonstrate valuing of clear, effective writing.

Three areas of particular weakness/need evident from reviewing the student work are as follows:

- Audience awareness – students often are not able to consider the needs of the reader or to adapt to likely readers.
- Mechanics – may not be the most important aspect of writing but errors in many papers are very distracting. This is NOT universal.
- Logic/rationale/critical thought – the higher order thinking that makes a written product more than a summary or report is often lacking.
Areas of strength include:

- Documentation – writers may or may not follow conventions (but these do vary by discipline and assignment), but they generally give credit where credit appears to be due.
- Sense of purpose – the papers generally include some sense of purpose which is often evident in the first part of the paper. Perhaps the purposes are assigned, at least in some cases, but students seem to be taking them to heart.
- Research design – students are able to conceptualize, carry out, and describe a project.

CONCLUDING COMMENTARY: WHERE DO WE STAND?

There is some reason for satisfaction with results for both critical thinking and written communication, but there is also evidence that action to create improvements in these key ES outcomes is needed. In terms of critical thinking, the scores suggest that faculty can feel satisfied that students are gaining competency in the various dimensions of this skill. At the same time, Findings suggest that more emphasis on higher order thinking skills, especially Evaluation & Conclusions, is warranted.

In written communication, findings show that most students can demonstrate competency in some important aspects of this skill, while still needing improvement in others. Since there is no clear pattern of strengths and weaknesses that is common across students generally, it is more difficult to say how that emphasis might be achieved.

In addition, it is important to note that senior students, in any numbers, who score at very low levels are a reason for concern. As students near graduation and complete capstone courses, we expect to see the best work they are capable of producing. When that “best work” is substantially below par, there is a basis for concern.

However, it is also worth noting that an assessment project of this sort is always partial and produces indications rather than facts. When faculty from across disciplines score together, it can be difficult for each person to put aside his or her own disciplinary perspectives despite the importance of norming and agreeing to consensus definitions and scores. When papers vary widely in length and type of assignment, expectations can be established by the first papers scored, and later papers may have strengths (or weaknesses) that are not fully noted in scoring because of assignment variability. Furthermore, the assignments responded to by some students may have required (e.g.) attention to some aspects of critical thinking or written communication but not others.

As additional scorings are completed over time, patterns will become clearer and more meaningful. In addition, data from other sources (student self-scorings/ratings, revalidation materials submitted for capstone courses, etc.) will help flesh out the picture.
Given these considerations, the findings represent a reason for attention but also evidence that some level of learning around critical thinking and written communication is occurring. More emphasis of those skills may allow a larger percentage of student work products to be scored in a range likely to be viewed as desirable for graduating seniors.

Attachments:

1. Table 1: Critical Thinking Results
2. Table 2: Written Communication Results
3. Figure 1: Rubric for Critical Thinking
4. Figure 2: Rubric for Written Communication