

BACHELOR OF INTERDISCIPLINARY STUDIES WRITING ASSESSMENT AUGUST 2009

In summer 2008, UVa's Office of Institutional Assessment and Studies (IAS) began coordinating the assessment of undergraduate writing competence University-wide. Recognizing the decentralized nature of curriculum and instruction at the University, individual schools were invited to design assessments of student writing that would address their own purposes and goals. The Bachelor of Interdisciplinary Studies (BIS) program proposed an assessment of student writing on liberal studies and capstone papers to investigate strengths and weaknesses in student writing both at the beginning (liberal studies) and the end of the program (capstone). Both sets of papers were assessed on the same learning outcomes, which were developed by BIS faculty. A descriptive scoring rubric was designed to assess student performance on the learning outcomes and BIS faculty used the rubric to evaluate the papers in two workshops that were facilitated by IAS. The results were tabulated by IAS, and this report presents those results, as well as a detailed description of the methodology. The rubric and details about inter-rater reliability follow in two appendices.

Methodology

In fall 2008, 27 capstone papers were collected from all graduating students in the BIS program in order to assess the writing ability of students *leaving* the program. In spring 2009, 46 liberal studies papers were selected from a larger pool of papers collected across several liberal studies courses to assess the writing ability of students *beginning* the program. This approach allows a cross-sectional view of a cohort of students entering the BIS program and a cohort leaving. The papers were assessed using a rubric that was developed by faculty in the BIS program with the assistance of IAS and Ryan Cordell, assistant director of the UVa Writing Center. The rubric outlined the following eight learning outcomes (full descriptions of the criteria for each learning outcome can be found in the copy of the rubric, Appendix A):

1. Introduction
2. Parts of Argument I – Claims and Subclaims
3. Parts of Argument II - Evidence
4. Counterarguments
5. Cohesion and Coherence I – Grammar and Style
6. Cohesion and Coherence II - Logic
7. Audience and Tone
8. Conclusions

The assessment of capstone papers was conducted in December 2008 with four BIS faculty raters and one English graduate student instructor rater. One paper was read and evaluated during the norming session in an effort to ensure that all raters were applying similar standards on all learning outcomes. The learning outcomes were rated on a scale of one to four, with a four representing the highest achievement and a one representing the lowest. Each performance level from one to four was defined with specific criteria. For example, for the Parts of Argument II - Evidence learning outcome, a score of one indicates that the writer's "claims are usually unsupported by evidence" and a score of four indicates that "claims are almost always supported by precise, authoritative, and varied evidence."

All student papers were de-identified to protect student confidentiality. Each paper was scored by two different raters. The assessment of liberal studies papers was conducted in May 2009 using the same four BIS faculty and English graduate student raters. The methodology for the norming session and the scale of rating in May was identical to the December session.

Rater bias was controlled by randomly assigning papers to raters and de-identifying all student papers. The raters were aware, however, that they were assessing a capstone or a liberal studies paper so rater bias in this respect could not be controlled. The research design is not “pre-post” as the papers in the two groups did not come from the same group of students. The cross-sectional design does allow a look at students’ writing abilities coming into the program and students leaving the program but the individual variations in student abilities within each group cannot be controlled in this design. The two groups’ performances will be compared to examine strengths and weaknesses in student achievement of the learning outcomes at the two different points in the program—entry and exit.

Inter-rater Reliability

Inter-rater reliability is a measure of how much consensus there is in the ratings made by different evaluators. The intra-class coefficient (ICC) was used to measure the reliability of the raters in this assessment because the ICC takes into account the differences in ratings for individual segments along with the overall correlation between raters. The ICC ranges from zero to one, with zero indicating little or no agreement and one indicating perfect agreement between raters. Overall, the inter-rater reliability was low, at 0.36 for the beginning-of-course papers and 0.30 for the end-of-course papers, indicating that raters did not often mark each individual paper with the *exact* same score for each learning outcome. The reliability fluctuated based on the learning outcomes, with some learning outcomes being rated more consistently than others (for a complete listing of reliabilities, see Appendix B). There was also a difference in rating style among the raters; some raters were more critical than others. Another reason for lower reliability was the high number of raters to papers (5 raters for 27 Capstone papers and 5 raters for 46 Liberal studies papers). The advantage to having more raters is that it increases faculty buy-in for, and experience with, assessment.

The raw agreement, however, presents a more easily interpreted view of the raters’ agreement. On average, approximately 45% of the ratings were exact matches between rater 1 and rater 2, 45% of ratings differed by only one point, and only 10% of ratings differed by more than one point. In order to correct for the ratings that differed by a point or more, all final scores on each learning outcome are the average score of both raters. Thus rater differences were reduced. Nevertheless, future assessments should include a more extended norming session, with more than one paper and perhaps an additional “mini” norming session midway through the process.

Results

The percentage of papers rated 1-4 for each learning outcome are presented in Tables 1 and 2. For ease of comparison, each learning outcome score (from 1 to 4) was standardized on a 0-100 point scale. An overall score, the average of all eight learning outcome scores, was also computed. The standardized mean scores for each group are presented in Table 3.

Table 1. Percentage of Ratings by Score for Each Learning Outcome - Liberal Studies

	1.00	2.00	3.00	4.00
Introduction	13.0%	64.0%	19.0%	4.0%
Parts of Argument I – Claims & Subclaims	24.0%	54.0%	19.0%	3.0%
Parts of Argument II - Evidence	5.0%	59.0%	32.0%	4.0%
Counterarguments	39.0%	39.0%	15.0%	7.0%
Cohesion & Coherence I – Grammar & Style	10.0%	48.0%	32.0%	10.0%
Cohesion & Coherence II - Logic	11.0%	56.0%	28.0%	5.0%
Audience & Tone	12.0%	62.0%	21.0%	5.0%
Conclusions	33.0%	42.0%	23.0%	2.0%

Table 2. Percentage of Ratings by Score for Each Learning Outcome - Capstone

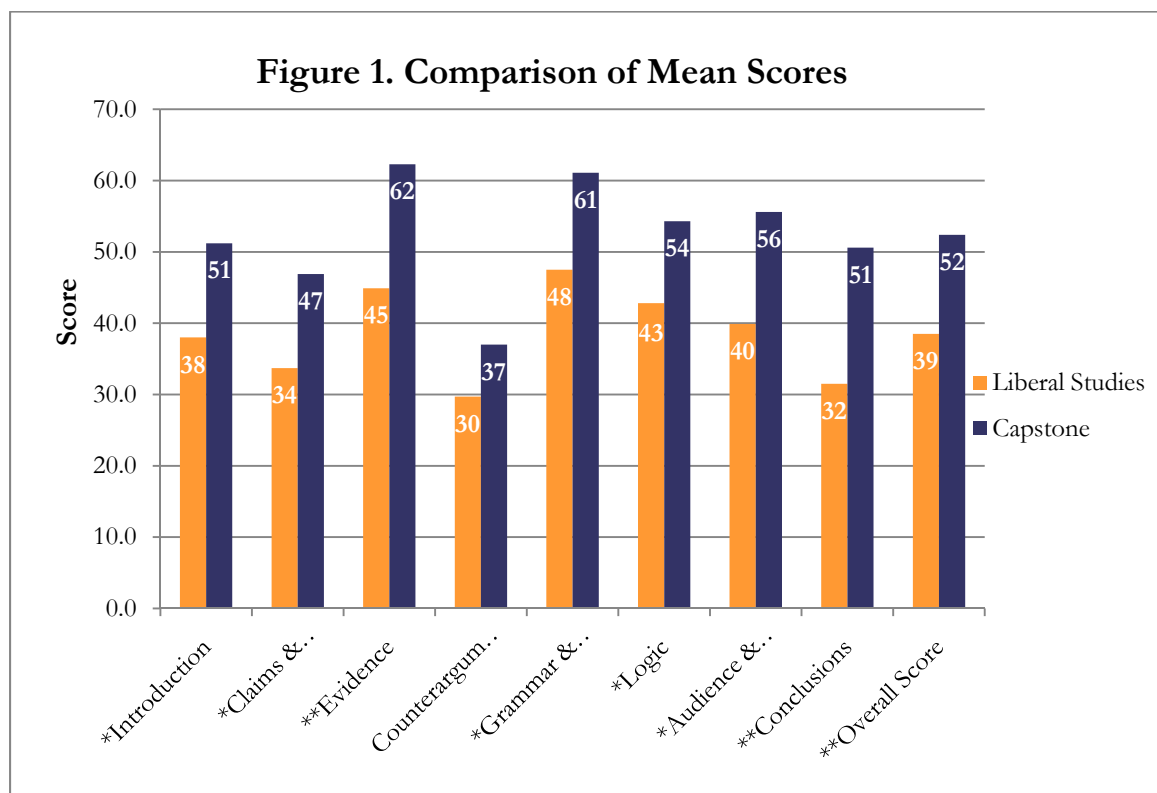
	1.00	2.00	3.00	4.00
Introduction	2.0%	57.0%	26.0%	15.0%
Parts of Argument I – Claims & Subclaims	6.0%	61.0%	20.0%	13.0%
Parts of Argument II - Evidence	0.0%	26.0%	61.0%	13.0%
Counterarguments	28.0%	41.0%	24.0%	7.0%
Cohesion & Coherence I – Grammar & Style	5.0%	24.0%	52.0%	19.0%
Cohesion & Coherence II - Logic	0.0%	59.0%	19.0%	22.0%
Audience & Tone	0.0%	52.0%	30.0%	18.0%
Conclusions	5.0%	54.0%	24.0%	17.0%

Table 3. Mean Scores on Learning Outcomes

	Liberal Studies Papers	Capstone Papers
Introduction*	38	51
Parts of Argument I – Claims & Subclaims*	34	47
Parts of Argument II - Evidence**	45	62
Counterarguments	30	37
Cohesion & Coherence I – Grammar & Style*	47	61
Cohesion & Coherence II - Logic*	43	54
Audience & Tone*	40	56
Conclusions**	32	51
Overall Score**	38	52

* Mean differences in performance are significant at $p < 0.05$.**Mean differences in performance are significant at $p < 0.001$.

The capstone group significantly outperformed the liberal studies group on all learning outcomes except for Counterarguments; performance on Counterarguments was statistically equal between the two groups. The greatest difference in performance was on the Conclusions and Parts of Argument II - Evidence learning outcomes. The smallest, but still significant, difference in performance was on Cohesion & Coherence II – Logic (see Figure 1).



*Mean differences in performance are significant at $p < 0.05$.

**Mean differences in performance are significant at $p < 0.001$.

Correlations among learning outcomes

In an effort to examine how performance on one learning outcome was correlated with performance on other learning outcomes, a correlation matrix comparing the relationships between each of the eight learning outcomes was created. The correlation matrix for the liberal studies group illustrates that with a few exceptions, performance on each learning outcome is significantly correlated with performance on all other learning outcomes (See Table 4). This is partly due to the fact that the students' scores in the liberal studies group are less varied (i.e. they are collectively lower) and more likely to be related. The one learning outcome that was most strongly correlated with other learning outcomes was Audience and Tone. Students' performance on Audience and Tone was significantly related to their performance on every other learning outcome. Evidence, Logic, and Conclusions were also significantly related to all other learning outcomes.

Table 4. Correlations among learning outcomes - Liberal Studies

	Introduction	Claims & Subclaims	Evidence	Counterarguments	Grammar & Style	Logic	Audience & Tone	Conclusions
Introduction	.	.61**	.32*	.30*	.16	.29*	.49**	.72**
Claims & Subclaims	.61**	.	.62**	.67**	.14	.49**	.70**	.73**
Evidence	.32*	.62**	.	.57**	.43**	.52**	.67**	.56**
Counterarguments	.30*	.66**	.57**	.	.14	.37*	.55**	.58**
Grammar & Style	.16	.14	.43**	.14	.	.51**	.46**	.37*
Logic	.29*	.49**	.52**	.37*	.51**	.	.61**	.44**
Audience & Tone	.49**	.70**	.67**	.55**	.46**	.61**	.	.65*
Conclusions	.73**	.73**	.56**	.58**	.34*	.44**	.65**	.

*Correlations are significant at $p < 0.05$.**Correlations are significant at $p < 0.01$.

For the end-of-course group, the correlations between learning outcomes were still strong among several learning outcomes (See Table 5). Performance on Introduction, Evidence, Logic, and Audience and Tone was significantly correlated with performance on all other learning outcomes. Performance on Claims and Subclaims, Counterarguments, and Conclusions was significantly correlated with performance on all other learning outcomes except one.

Table 5. Correlations among learning outcomes - Capstone

	Introduction	Claims & Subclaims	Evidence	Counterarguments	Grammar & Style	Logic	Audience & Tone	Conclusions
Introduction	.	.77**	.42*	.74**	.44*	.60**	.62**	.71**
Claims & Subclaims	.77**	.	.68**	.79**	.37	.72**	.53**	.75**
Evidence	.42*	.68**	.	.55**	.48*	.67**	.53**	.54**
Counterarguments	.74**	.79**	.55**	.	.31	.60**	.58**	.59**
Grammar & Style	.44*	.37	.48*	.31	.	.43*	.65**	.22
Logic	.60**	.72**	.67**	.60**	.43*	.	.71**	.71**
Audience & Tone	.62**	.53**	.53**	.58**	.65**	.71**	.	.52**
Conclusions	.71**	.75**	.54**	.59**	.22	.71**	.52**	.

*Correlations are significant at $p < 0.05$.**Correlations are significant at $p < 0.01$.

Appendix A

BIS Writing Assessment Rubric

Introduction

4 - Introduction clearly explains and proposes to solve a problematic attitude, idea, or practice in the writer's field or subject of study. The proposed solution (the thesis or claim) is innovative, compelling, and convincing, motivating reader interest in both the paper and the larger field or subject.

3 - Introduction explains and proposes to solve a problematic attitude, idea, or practice in the writer's field or subject of study. The proposed solution (the thesis or claim) motivates reader interest in the paper.

2 - Introduction gestures toward a problematic attitude, idea, or practice in the writer's field or subject of study, but the problem is implicit rather than explicitly stated. Proposed solution (the thesis or claim) is unclear, unfocused, or too simplistic to satisfactorily address the problem.

1 - Introduction demonstrates confusion or misunderstanding about the writer's field or subject of study. Introduction provides no context for readers, is vague, and proposes no solutions (no thesis or claim).

Parts of Argument I - Claims and Subclaims

4 - Argument is well balanced, with specific, insightful, debatable claims and sub-claims. Ideas progress in a logical sequence, work to support a clear structure, and claims are placed at appropriate intervals.

3 - Argument is balanced, with specific claims. Ideas progress in a logical sequence and claims are placed at appropriate intervals.

2 - Writers argument is unwieldy; some claims are implicit, unclear, or nearly indisputable. Makes some generalizations without support.

1 - Argument is unbalanced or impossible to identify. Claims are either entirely unclear or indisputable. Frequently makes illogical generalizations without support.

Parts of Argument II - Evidence

4 - Claims are almost always supported by precise, authoritative, and varied evidence.

3 - Claims are usually supported by authoritative and varied evidence.

2 - Evidence is sometimes insufficient, unreliable (unauthorized sources, anecdote, etc.), or only loosely connected to claims.

1 - Claims are usually unsupported by evidence.

Counterarguments

- 4 - Acknowledges and sufficiently explains counterarguments, responding thoroughly and convincingly through dialogue rather than verbal combat.
- 3 - Acknowledges and explains counterarguments, responding through dialogue rather than verbal combat.
- 2 - Acknowledges counterarguments, but does not fully explain them. Responds to counterarguments in a cursory manner, or combatively rather than dialogically.
- 1 - Either does not acknowledge any counterarguments or responds to them in an excessively combative or hostile manner.

Cohesion and Coherence I - Grammar and Style

- 4 - Writing is cohesive and coherent: information flow within sentences and paragraphs is logical and consistent. Sentences are strong, expressive, and varied in construction. Prose is stylistic and compelling. Grammar, diction, and spelling are nearly perfect.
- 3 - Writing is cohesive and coherent: information flow within sentences and paragraphs is logical and consistent. Sentences are clear, but may be formulaic or tedious. Document contains some common errors in grammar, diction, and spelling.
- 2 - Writing is fragmented: information flow between sentences and paragraphs is inconsistent. Sentences demonstrate little or no variety in style.
- 1 - Syntax may be irregular, and the document may contain persistent errors in grammar, diction, and spelling that hamper meaning. Writing is incoherent and fragmented. Problems with syntax create barriers to reader understanding, and the document contains pervasive errors in grammar, diction, and spelling.

Cohesion and Coherence II - Logic

- 4 - Project as a whole demonstrates a logically coherent structure that is clearly conveyed to the reader.
- 3 - Project is logically coherent, but this coherence could be more clearly articulated.
- 2 - Sections of the project are logically coherent.
- 1 - Project demonstrates no discernible logical coherence.

Audience and Tone

- 4 - Exhibits a thorough understanding of the goals, readers, situation, purpose, and structure of their argument, and writes in a style appropriate to each. Authorial tone is consistent, mature, and engaging. The language appropriately academic.
- 3 - Exhibits a solid understanding of the goals, readers, situation, purpose, and structure of their argument. Authorial tone is consistent, though perhaps unrefined or static. Language is appropriately academic.
- 2 - Exhibits an inconsistent understanding of the goals, readers, situation, purpose, and structure of their argument. Authorial tone is uneven or immature. Language is occasionally nonacademic.
- 1 - Exhibits little or no understanding of the goals, readers, situation, purpose, and structure of their argument. Argument is severely hampered by pervasive stylistic problems. Authorial tone is unbalanced and immature. Language is frequently nonacademic and distracts from the argument.

Conclusions

- 4 - Conclusions are logical, convincing, clearly expressed, and consistent with those proposed in the introduction.
- 3 - Conclusions are logical and consistent with those proposed in the introduction.
- 2 - Conclusions are logical, but either unconvincing, not clearly expressed, or inconsistent with those promised in the introduction.
- 1 - Conclusions are not drawn, indecipherable, illogical, or inconsistent with the information presented in the rest of the document.

Appendix B

Table B.1 Inter-Rater Reliability

	Liberal Studies	Capstone
Introduction	0.32	0.50
Parts of Argument I – Claims & Sub-claims*	0.22	0.36
Parts of Argument II - Evidence	0.37	0.11
Counterarguments	0.57	0.48
Cohesion & Coherence I – Grammar & Style	0.38	0.52
Cohesion & Coherence II - Logic	0.24	0.01
Audience & Tone	0.52	0.12
Conclusions**	0.24	0.26