

UNIVERSITY OF VIRGINIA
Jonathan Schnyer, Associate Director,
Office of Institutional Assessment and Studies (IAS)

WRITING 2008
PART B: STATUS REPORT

4.0 DATA PRESENTATION

- 4.1 Getting students to respond to requests for voluntary participation in assessments is always a challenge. Although we were able to collect a representative sample of 179 first-year student papers from ENWR courses, we had to invite representative samples of fourth-year students in the College of Arts and Sciences and the Schools of Architecture and Nursing to voluntarily submit their best work by email. One of the main challenges was ensuring that we received a sample of papers that was representative of all disciplines. Social Science and Science and Math majors, in particular, were targeted with specific reminder emails because they were underrepresented groups. In our final sample, Humanities and Fine Arts majors are slightly overrepresented and Social Science and Architecture majors are slightly underrepresented (See Table 1).

Table 1. Representativeness of the Fourth-Year Sample

	Sample	Population
College of Arts & Sciences Humanities and Fine Arts	42.8%	36.3%
College of Arts & Sciences Science and Math	25.6%	26.4%
College of Arts & Sciences Social Science	27.0%	31.2%
Architecture	2.5%	3.4%
Nursing	2.0%	2.7%
Total N	355	3122

Ratings and Reliability. We had to forgo the use of a third rating this year because the cost was too high. All final scores on each learning outcome are the average of two ratings. The overall reliability of the ratings ranged from 0.41 to 0.61 with an average reliability that was moderate-low at 0.54. One reason for the lower reliability is the high number of raters (approximately 13 raters for each assessment workshop) and the fact that this was a first-time assessment using the rubric. The raw agreement, however, presents a more easily interpreted view of reliability. On average, 49% of the ratings were exact matches between rater one and rater two, 44% differed by one point, and only 7% of the ratings differed by more than one point.

- 4.2 **Cross-Sectional, Value-Added Results.** All student papers were assessed using the same skills-based descriptive scoring rubric developed by faculty in the English department with particular help provided by the assistant director of the UVa Writing Center and IAS. The eight learning outcomes on the writing rubric include the following: Introduction, Parts of Argument I – Claims and

Subclaims, Parts of Argument II – Evidence, Counterarguments, Cohesion and Coherence, Audience and Tone, and Conclusions (see Appendix A for a full copy of the rubric and the descriptive scoring criteria for each learning outcome). The individual skills were assigned a score of 4 (highly competent), 3 (competent), 2 (minimally competent) or 1 (not competent); and an overall score for each student was calculated by summing up and averaging the scores for each individual skill.

Fourth-year students in the College of Arts and Sciences (CLAS) and the Schools of Architecture and Nursing significantly outperformed first-year students overall and on every individual writing learning outcome ($p < .001$) (See Table 2).

Table 2. Mean Differences Between Groups

	1st-Years Mean Score	4th-Years Mean Score
Introduction*	2.03	2.52
Parts of Argument I- Claims and Subclaims*	1.98	2.68
Parts of Argument II- Evidence*	2.04	2.82
Counterarguments*	1.73	1.97
Cohesion and Coherence*	2.25	2.93
Audience and Tone*	2.04	2.85
Conclusions*	2.05	2.75
Final Score*	2.02	2.64
N	179	355

*Mean difference is significant at $p < .001$

School-Specific Value-Added Results. The School of Engineering and Applied Science (SEAS), the McIntire School of Commerce, and the Bachelor of Interdisciplinary Studies program (BIS) all designed individual assessments that addressed their own purposes and goals with respect to student writing. The School of Engineering and Applied Science (SEAS) assessed the writing of students both entering and graduating from the program. First-year student work was randomly selected from the first writing course in the Department of Science, Technology, and Society (STS 101) and fourth-year student work was randomly selected from a portion of the students' thesis papers written in their STS 402 course. The two groups of student papers were assessed using the same skills-based descriptive rubric in two assessment workshops. Fourth-year students outperformed first-year students overall and on four individual learning outcomes. The report of SEAS results can be found in Appendix B.

The McIntire School of Commerce assessed the memorandum writing of third-year students at the beginning of the Integrated Core Experience (ICE) and after the students completed the first series of ICE courses (i.e. beginning of fall semester and end of fall semester). Forty papers were randomly selected at the beginning and another 40 papers were selected at the end of the semester. These papers were assessed using the same skills-based descriptive scoring rubric in an

assessment workshop. End-of-course student papers received significantly higher scores overall and on seven individual learning outcomes. The report of Commerce results can be found in Appendix C.

The Bachelor of Interdisciplinary Studies Program (BIS) in the School of Continuing and Professional Studies assessed the writing of students both entering and graduating from the program. Student work was randomly selected from several introductory liberal studies courses and compared to Capstone papers which were collected from graduating students. Student papers were assessed using the same skills-based descriptive scoring rubric in two assessment workshops. The graduating students outperformed the entering students overall and on all writing learning outcomes. The report of BIS results can be found in Appendix D.

Competence. The competency standards that were established for graduating fourth years were: “40% of undergraduates are expected to be highly competent; 85% of undergraduates competent or above; and 100% minimally competent or above.” Minimal competence was defined as an overall score of 1.50-2.49, competence was defined as an overall score of 2.50-3.49, and high competence was defined as an overall score equal to or greater than 3.50. Fourth-year students almost met the competency standard for minimally competent but did not reach the standard for competent (85%) or highly competent (40%). Given that this was the first administration of the assessment, the standards need to be re-examined now that real data can inform the discussion of the expected standards (See Tables 3 and 4).

Table 3. Fourth-Year Writing Competency (All Schools*)

	Minimally Competent or above	Competent or above	Highly Competent
All Fourth-Years	99%	61%	8%
Standards for Competency (see Section 1.2)	100%	85%	40%
N	514	317	42

*Data is weighted by school.

Table 4. Fourth-Year Writing Competency (CLAS, Architecture, Nursing)

	Minimally Competent or above	Competent or above	Highly Competent
All Fourth-Years	99%	61%	8%
Standards for Competency (see Section 1.2)	100%	85%	40%
N	351	216	28

The competency results were also broken down by discipline within the College of Arts and Sciences. The results were not broken down by Architecture or Nursing because the sample sizes of these groups were too low to make the analysis meaningful. The mean scores of the three disciplines were not

significantly different, but the Humanities and Fine Arts and Social Science disciplines had more student papers scored as competent than the Science and Math discipline (See Table 4).

Table 5. Fourth-Year Writing Competency by Discipline within the College of Arts and Sciences

	Minimally Competent or above	Competent or above	Highly Competent	N
Humanities and Fine Arts	98%	64%	8%	152
Science and Math	100%	51%	9%	91
Social Science	99%	67%	7%	96

- 4.3 After careful consideration of the results, the University may seek additional data from faculty or students about the writing skills of its students, particularly the scientific writing skills of Science and Math majors. Other follow-up measures could take the form of surveys, focus groups, or additional assessments.

APPENDIX A

COLLEGE OF ARTS AND SCIENCES AND SCHOOLS OF ARCHITECTURE AND NURSING 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 - Writer's 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

- 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. Syntax may be irregular, and the document may contain persistent errors in grammar, diction, and spelling that hamper meaning.
- 1 - Writing is incoherent and fragmented. Problems with syntax create barriers to reader understanding, and the document contains pervasive errors in grammar, diction, and spelling.

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

SCHOOL OF ENGINEERING AND APPLIED SCIENCES 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 School of Engineering and Applied Sciences (SEAS) proposed an assessment of student writing on Science, Technology and Society papers to investigate strengths and weaknesses in student writing both at the beginning (STS 101) and the end of the program (STS 402). Both sets of papers were assessed on the same learning outcomes, which were developed by SEAS faculty. A descriptive scoring rubric was designed to assess student performance on the learning outcomes, and SEAS faculty used the rubric to evaluate the papers in two workshops that were facilitated by IAS. In addition to the in-house rubric devised by SEAS faculty, the end-of-program papers were rated on four learning objectives from the Accreditation Board for Engineering and Technology (ABET). 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, a random sample of 101 STS 101 papers was collected from all entering students in the School of Engineering and Applied Sciences in order to assess the writing ability of students *beginning* the program. In spring 2009, a random sample of 97 STS 402 papers was collected from all graduating students to assess the writing ability of students *leaving* the program. This approach allows a cross-sectional view of a cohort of students entering the SEAS program and a cohort leaving. All papers were assessed using a rubric that was developed by faculty in SEAS with the assistance of IAS and Ryan Cordell, assistant director of the UVa Writing Center. The rubric outlines the following six learning outcomes (full descriptions of the criteria for each learning outcome can be found in the copy of the rubric, Appendix B.1):

1. Introduction
2. Audience Occasion and Purpose
3. Cohesion and Coherence
4. Grammar
5. Graphics
6. Conclusions

In addition, the STS 402 papers were rated on four learning objectives from ABET:

1. Impacting global and societal context
2. Integration of contextual issues
3. Professional and ethical responsibility
4. Role of engineering in contemporary issues

The assessment of STS 101 papers was conducted in December 2008 with eight SEAS faculty raters and one English graduate student instructor rater. Two papers were read and evaluated during a 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 Audience Occasion and Purpose learning outcome, a score of one indicates that the writer “exhibits little or no understanding of the goals, readers, situation, purpose, and structure of the argument” and a score of four indicates that the writer “exhibits a thorough understanding of the goals, readers, situation, purpose, and structure of the argument, and writes in a style appropriate to each.” All student papers were de-identified to protect student confidentiality. Each paper was scored twice by two different raters.

The assessment of STS 402 papers was conducted in May 2009 with six SEAS faculty raters and three English graduate student raters. The methodology for the norming session and the scale of rating 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 STS 101 or a STS 402 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’ performance will be compared to examine strengths and weaknesses in students’ 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 0.24 for the beginning-of-course papers and 0.19 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.2). 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 (9 raters for 101 STS 101 papers and 9 raters for 97 STS 402 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. For the set of six learning outcomes that are directly compared between STS 101 and STS 402, approximately 47% of the ratings, on average, were exact matches between rater 1 and rater 2, 40% of ratings differed by only one point, and only 13% of ratings differed by more than one point. For the four ABET learning objectives, raters were an exact match, on average, 38% of the time, they were one point apart 48% of the time and differed by more than one point 14% of the time. 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 two papers, 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 mean scores for each group are presented in Table 3.

Table 1. Percentage of Ratings by Score for Each Learning Outcome - STS 101

	1	2	3	4
Introduction	6%	42%	34%	18%
Audience Occasion and Purpose	6%	43%	44%	7%
Cohesion and Coherence	4%	41%	48%	7%
Grammar	5%	31%	46%	18%
Graphics	1%	0%	0%	99%
Conclusions	14%	46%	28%	12%

Table 2. Percentage of Ratings by Score for Each Learning Outcome - STS 402

	1	2	3	4
Introduction	6%	40%	34%	20%
Audience Occasion and Purpose	3%	37%	46%	14%
Cohesion and Coherence	1%	23%	57%	19%
Grammar	0%	13%	48%	39%
Graphics	8%	6%	27%	59%
Conclusions	4%	45%	37%	14%

Table 3. Mean Scores on Learning Outcomes

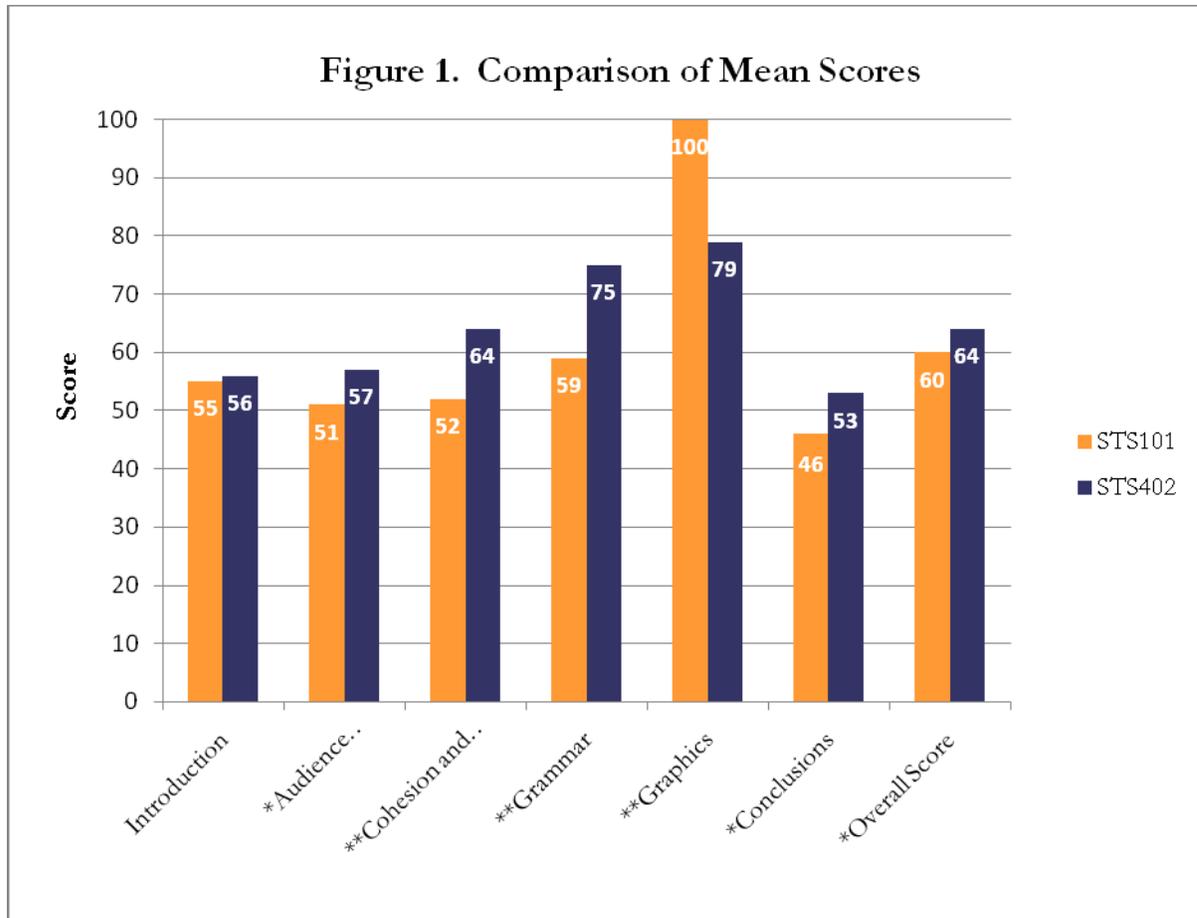
	STS 101	STS 402
Introduction	55	56
Audience Occasion and Purpose*	51	57
Cohesion and Coherence**	52	64
Grammar**	59	75
Graphics**	100	79
Conclusions*	46	53
Overall Score*	60	64

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

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

The STS 402 group significantly outperformed the STS 101 group on all learning outcomes except for Introduction; performance on Introduction was statistically equal between the two groups.¹ The greatest difference in performance was on the Grammar and Cohesion and Coherence learning outcomes. The smallest, but still significant, difference in performance was on Audience Occasion and Purpose (see Figure 1).

¹ The STS 101 group outperformed the STS 402 group on the graphics learning outcome only because the STS 101 group did NOT present graphics in their papers and were not expected to; thus, they received the full amount of points on that learning outcome. The STS 402 students, however, were expected to use graphics and were assessed on them, increasing the likelihood that they would get a lower score.



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

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

The STS 402 group was also evaluated on the learning outcomes from the Accreditation Board for Engineering and Technology (ABET), which certifies Engineering programs across the United States. The mean scores for each group are presented in Table 4.

Table 4. Mean Scores on ABET Learning Outcomes (STS 402 Only)

	STS 402
Impacting global and societal context	57
Integration of contextual issues	50
Professional and ethical responsibility	44
Role of engineering in contemporary issues	53
Overall Score	51

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 relationship between each of the six learning outcomes was created. The correlation matrix for the STS 101 group illustrates that with a few exceptions, performance on each learning outcome is

significantly correlated with performance on all other learning outcomes (See Table 5).² Two learning outcomes were highly correlated with the overall score: Cohesion and Coherence, and Audience Occasion and Purpose.

Table 5. Correlations among learning outcomes – STS 101

	Introduction	Audience Occasion & Purpose	Cohesion & Coherence	Grammar	Graphics	Conclusions
Introduction	.	.42**	.55**	.33**	-.22*	.46**
Audience Occasion & Purpose	.42**	.	.62**	.54**	-.09	.65**
Cohesion & Coherence	.55**	.62**	.	.59**	-.08	.59**
Grammar	.33**	.54**	.59**	.	-.12	.40**
Graphics	-.22*	-.09	-.08	-.12	.	-.16
Conclusions	.46**	.65**	.59**	.40**	-.16	.

* Correlation is significant at $p < 0.05$.

** Correlation is significant at $p < 0.01$.

For the end-of-course group, the correlations between learning outcomes were also very strong among several outcomes, especially between Conclusions and Audience and Purpose, and between Grammar and Cohesion and Coherence (See Table 6). Conclusions and Audience Occasion and Purpose were the two learning outcomes that correlated most strongly with the Overall Score.

Table 6. Correlations among learning outcomes – STS 402

	Introduction	Audience Occasion & Purpose	Cohesion & Coherence	Grammar	Graphics	Conclusions
Introduction	.	.67**	.45**	.18	.28**	.54**
Audience Occasion & Purpose	.67**	.	.51**	.34**	.17	.58**
Cohesion & Coherence	.45**	.51**	.	.56**	.20*	.47**
Grammar	.18	.34**	.56**	.	.18	.35**
Graphics	.28**	.17	.20*	.18	.	.31**
Conclusions	.54**	.58**	.47**	.35**	.31**	.

* Correlation is significant at $p < 0.05$.

** Correlation is significant at $p < 0.01$.

² Negative correlation coefficients for the graphics learning outcome are explained by the fact that all students obtained the same score on this outcome because there was no requirement for graphics in the STS 101 paper. Thus, in this case, these coefficients are meaningless.

Correlations among ABET learning outcomes were all quite strong (See Table 7).

Table 7. Correlations among learning outcomes - ABET

	Impacting global and societal context	Integration of contextual issues	Professional and ethical responsibility	Role of engineering in contemporary issues
Impacting global and societal context	.	.67**	.43**	.70**
Integration of contextual issues	.67**	.	.49**	.68**
Professional and ethical responsibility	.43**	.49**	.	.47**
Role of engineering in contemporary issues	.70**	.68**	.47**	.

* Correlation is significant at $p < 0.05$.

** Correlation is significant at $p < 0.01$.

APPENDIX B.1

SEAS WRITING ASSESSMENT RUBRIC

Introduction

- 4 – Introduction quickly, reliably, and clearly establishes context and explicitly signals document’s purpose to its reader.
- 3 – Introduction reliably establishes context, but may be overlong or tedious. The introduction signals the document’s purpose to readers.
- 2 – Introduction implicitly signals the document’s purpose to readers, but is excessively brief and does not fully establish the context of the paper for the reader.
- 1 – Introduction is either missing, unacceptably brief, or entirely unreliable, with no connection to the body of the document.

Audience Occasion and Purpose

- 4 – Exhibits a thorough understanding of the goals, readers, situation, purpose, and structure of the 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 the 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 the 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 the argument. Argument is severely hampered by pervasive stylistic problems. Authorial tone is unbalanced and immature. Language is frequently nonacademic and distracts from the argument.

Cohesion and Coherence

- 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.
- 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.
- 2 – Writing is fragmented: information flow between paragraphs is consistent. Sentences demonstrate little or no variety in style. Syntax may be irregular.
- 1 – Writing is incoherent and fragmented: information flow between sentences and paragraphs is inconsistent. Problems with syntax create barriers to reader understanding.

Grammar

- 4 – Writer establishes credibility with nearly perfect grammar, diction, and spelling.
- 3 – Document contains some errors in grammar, diction, and spelling, but none that challenge reader understanding.
- 2 – Document contains several errors in grammar, diction, and spelling that begin to hurt the writer’s credibility.
- 1 – Document contains frequent or pervasive errors in grammar, diction, and spelling that create barriers to reader understanding and seriously hurt the writer’s credibility.

Graphics

- 4 – Writer integrates graphics where appropriate into the text or writer appropriately chooses not to integrate graphics into the document.
- 3 – Writer fails to include graphics where they would have enhanced readability, but were not absolutely necessary.
- 2 – Writer integrates graphics into the text inappropriately.
- 1 – Writer fails to integrate needed graphics into the document.

Conclusions

- 4 – Writer delineates methods, presents results, and reaches conclusions that are logical, clear, and consistent with the rest of the document.
- 3 – Writer delineates methods, presents results and reaches conclusions that are logical and consistent with the rest of the document.
- 2 – Writer attempts to delineate their methods and present results, and also gestures towards conclusions, but these conclusions are not fully or clearly expressed. Logic of the conclusions may be difficult for readers to follow.
- 1 – Writer either makes no attempt to delineate methods and present results, or reaches conclusions that are illogical or inconsistent with the information presented in the body of the document.

ABET RUBRIC FOR END-OF-PROGRAM ASSESSMENT

Professional and ethical responsibility

- 4 – Exhibits a thorough understanding of professional and ethical responsibility.
- 3 – Exhibits a solid understanding of professional and ethical responsibility.
- 2 – Exhibits an inconsistent understanding of professional and ethical responsibility.
- 1 – Exhibits little or no understanding of professional and ethical responsibility.

Impact in global and societal context

- 4 – Exhibits a thorough understanding of the impact of engineering solutions in global and societal context.
- 3 – Exhibits a solid understanding of the impact of engineering solutions in global and societal context.
- 2 – Exhibits an inconsistent understanding of the impact of engineering solutions in global and societal context.
- 1 – Exhibits little or no understanding of the impact of engineering solutions in global and societal context.

Role of engineering in contemporary issues

- 4 – Exhibits a thorough understanding of the role of technology and engineering in contemporary issues/use of historical context to put contemporary issues in perspective.
- 3 – Exhibits a solid understanding of the role of technology and engineering in contemporary issues/use of historical context to put contemporary issues in perspective.
- 2 – Exhibits an inconsistent understanding of the role of technology and engineering in contemporary issues/use of historical context to put contemporary issues in perspective.
- 1 – Exhibits little or no understanding of the role of technology and engineering in contemporary issues/use of historical context to put contemporary issues in perspective.

Integration of contextual issues

- 4 – Exhibits a thorough understanding of and thoroughly considers and integrates contextual issues into the design, implementation, and management of technological systems.
- 3 – Exhibits a solid understanding of and solidly considers and integrates contextual issues into the design, implementation, and management of technological systems.
- 2 – Exhibits an inconsistent understanding of and inconsistently considers and integrates contextual issues into the design, implementation, and management of technological systems.
- 1 – Exhibits little or no understanding, consideration, or integration of contextual issues into the design, implementation, and management of technological systems.

APPENDIX B.2**SEAS INTER-RATER RELIABILITY**

Table B.1. Inter-Rater Reliability In-House Rubric

	STS 101	STS 402
Introduction	0.17	0.16
Audience Occasion and Purpose	0.29	0.39
Cohesion and Coherence	0.28	0.25
Grammar	0.32	-0.11
Graphics	0.00	-0.03
Conclusions	0.41	0.36

Table B.2. Inter-Rater Reliability ABET

	STS 402
Impacting global and societal context	0.18
Integration of contextual issues	0.38
Professional and ethical responsibility	0.41
Role of engineering in contemporary issues	0.18

APPENDIX C

MCINTIRE SCHOOL OF COMMERCE 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 McIntire School of Commerce proposed an assessment of students' memorandum writing to investigate strengths and weaknesses in student writing both at the beginning and end of the fall semester in the students' third year. Both sets of student papers were assessed on the same learning outcomes which were developed by Commerce faculty. A descriptive scoring rubric was designed to assess student performance on the learning outcomes, and Commerce 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 September and December 2008, memorandum writing assignments were collected from all third-year Commerce students in the Integrated Core Experience. From these, 40 papers were randomly selected from the beginning-of-course group and another 40 were randomly selected from the end-of-course group. This approach allows a cross-sectional view of a cohort of students at the beginning and end of their third-year fall semester. The papers were assessed using a rubric that was developed by faculty in the School of Commerce 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 C.1):

1. Skim Value
2. Clear Purpose
3. Document Preview
4. Bottom Line Up Front (BLUF)
5. Cohesion and Coherence
6. Grammar
7. "Plain English" Style
8. Reader Expectations

The assessment was conducted in January 2009 with two Commerce faculty raters and four English graduate student instructor raters. Two papers were 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 Document Preview learning outcome, a score of one indicates that the "writer omits any mention of the document's structure" and a score of four indicates that the "writer explicitly previews the document's structure."

All student papers were de-identified to protect student confidentiality. Each paper was scored by two different raters. All papers were rated in one session in January 2009. Rater bias was controlled by randomly assigning papers to raters and de-identifying all student papers. Raters were also blind to whether they were rating an end-of-course or beginning-of-course paper. The research design is not a true pre-post design, since the 40 student papers from September were not matched with the same student papers from December. The cross-sectional design does allow a look at students' writing abilities at the beginning of the semester and at the end but individual variations in 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 two different points in the program—the beginning of the third-year fall semester and the end of the third-year fall semester.

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 for the beginning-of-course papers was moderate at 0.50 and low-moderate at 0.38 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 C.2). There was also a difference in rating style among the raters; some raters were more critical than others. Other reasons for lower reliability include the high number of raters to papers (6 raters for 40 papers) and the fact that this was a first-time assessment using the rubric to assess memorandum writing. The advantage to having more raters is that it increases faculty buy-in for, and experience with, assessment.

The raw agreement among raters presents a more easily interpreted view of the raters' agreement. On average, approximately 44% of the ratings were exact matches between rater 1 and rater 2, 50% of ratings differed by only one point, and only 6% 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 assessment should include a more extended norming session, with more than two papers 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. Each learning outcome score (from 1 to 4) was standardized on a 0-100 point scale for ease of comparison. 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 - Beginning of Course

	1	2	3	4
Skim Value	25%	36%	29%	10%
Clear Purpose	27%	48%	16%	9%
Document Preview	51%	31%	13%	5%
Bottom Line Up Front (BLUF)	29%	44%	21%	6%
Cohesion & Coherence	0%	31%	59%	10%
Grammar	1%	10%	46%	43%
Plain English Style	1%	24%	46%	29%
Reader Expectations	19%	40%	26%	15%

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

	1	2	3	4
Skim Value	0%	28%	41%	31%
Clear Purpose	4%	21%	36%	39%
Document Preview	21%	28%	34%	17%
Bottom Line Up Front (BLUF)	13%	48%	26%	13%
Cohesion & Coherence	0%	21%	49%	30%
Grammar	3%	14%	43%	40%
Plain English Style	0%	4%	26%	70%
Reader Expectations	0%	18%	41%	41%

Table 3. Mean Scores on Learning Outcomes

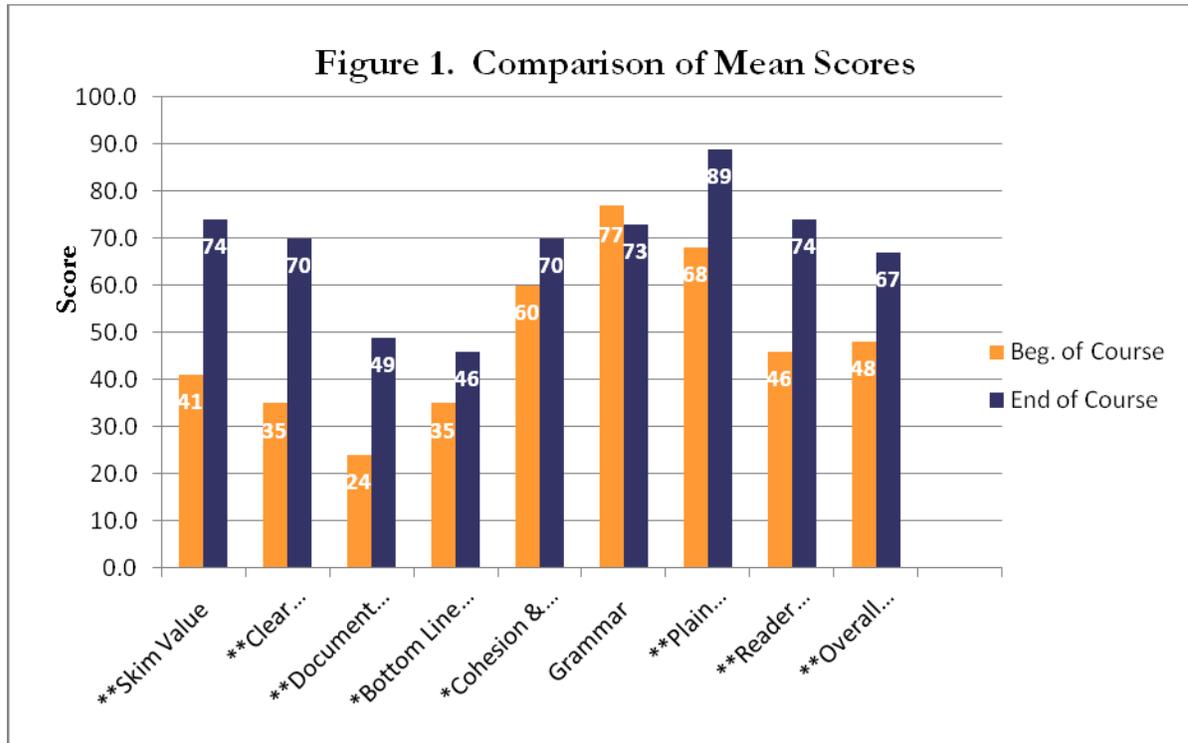
	Beginning of Course	End of Course
Skim Value**	41	74
Clear Purpose**	35	70
Document Preview**	24	49
Bottom Line Up Front (BLUF)*	35	46
Cohesion & Coherence*	60	70
Grammar	77	73
Plain English Style**	68	89
Reader Expectations**	46	74
Overall Score**	48	67

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

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

The end-of-course group significantly outperformed the beginning-of-course group on all learning outcomes except for Grammar; performance on Grammar was statistically equal between the two groups. The greatest difference in performance was on the Clear Purpose and Reader Expectations learning outcomes. Although significant differences exist on the Bottom Line Up Front and Document Preview learning outcomes, the mean performance of students

on these learning outcomes was lower in comparison to performance on the other six outcomes (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 another learning outcome, a correlation matrix comparing the relationships between each of the eight learning outcomes was created. The correlation matrix for the beginning-of-course 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 beginning-of-course 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 the other learning outcomes was Reader Expectations. Students' performance on Reader Expectations was significantly related to their performance on every other learning outcome. Document Preview, Skim Value, and Plain English Style also had strong correlations with at least five other learning outcomes.

Table 4. Correlations among learning outcomes - Beginning of Course

	Skim Value	Clear Purpose	Document Preview	BLUF	Cohesion & Coherence	Grammar	Plain English Style	Reader Expectations
Skim Value	.	.58**	.56**	.67**	.23	.25	.46**	.68**
Clear Purpose	.58**	.	.41**	.34*	.49**	.29	.49**	.75**
Document Preview	.56**	.41*	.	.43**	.32*	.21	.52**	.53**
BLUF	.67**	.34*	.43**	.	.29	.34*	.31	.54**
Cohesion & Coherence	.23	.49*	.32*	.29	.	.37*	.44**	.51**
Grammar	.25	.29	.21	.34*	.37*	.	.19	.40*
Plain English Style	.46**	.49**	.52**	.31	.44**	.19	.	.44**
Reader Expectations	.68**	.75**	.53**	.54**	.51**	.40*	.44**	.

* Correlation is significant at $p < 0.05$.** Correlation is significant at $p < 0.01$.

For the end-of-course group, the correlations between learning outcomes were not as numerous or as strong as those for the beginning-of-course group (See Table 5). Performance on Plain English Style was significantly correlated with performance on all other learning outcomes except one. Clear Purpose and Document Preview also had significant correlations with at least four other learning outcomes.

Table 5. Correlations among learning outcomes - End of Course

	Skim Value	Clear Purpose	Document Preview	BLUF	Cohesion & Coherence	Grammar	Plain English Style	Reader Expectations
Skim Value	.	.14	.42**	.19	.27	.31	.40*	.07
Clear Purpose	.14	.	.37*	.42**	.31	.00	.36*	.48**
Document Preview	.42**	.37*	.	.14	.16	.34*	.48**	.01
BLUF	.19	.42**	.14	.	.20	.02	.31	.20
Cohesion & Coherence	.27	.31	.16	.20	.	.42**	.36*	.34*
Grammar	.31	.00	.34*	.02	.42**	.	.48**	.02
Plain English Style	.40*	.36*	.48**	.31	.36*	.48**	.	.34*
Reader Expectations	.07	.48**	.01	.48**	.34*	.02	.34*	.

* Correlation is significant at $p < 0.05$.** Correlation is significant at $p < 0.01$.

APPENDIX C.1

MCINTIRE WRITING ASSESSMENT RUBRIC

Skim Value

- 4 – Writer visually guides the reader to all of the important elements in the document (high skim value).
- 3 – Writer visually guides the reader to most of the important elements in the document (moderate skim value).
- 2 – Writer visually guides the reader to some of the important elements in the document (low skim value).
- 1 – Writer makes little or no effort to visually guide the reader.

Clear Purpose

- 4 – Writer succinctly but fully delineates the occasion and purpose(s) for writing.
- 3 – Writer establishes the occasion and purpose(s) for writing.
- 2 – Writer gestures implicitly, rather than explicitly, toward the occasion and purpose(s) for writing.
- 1 – Writer makes no attempt to establish the occasion or purpose for writing.

Document Preview

- 4 – Writer explicitly previews the document's structure.
- 3 – Writer previews the document's structure.
- 2 – Writer inadequately hints at the document's structure.
- 1 – Writer omits any mention of the document's structure.

Bottom Line Up Front (BLUF)

- 4 – Writer presents all major conclusions and/or recommendations near the document's opening (Bottom Line Up Front/BLUF). Presentation is clear and well organized.
- 3 – Writer presents most major conclusions and/or recommendations near the document's opening. Presentation is clear.
- 2 – Writer gestures toward presenting major conclusions and/or recommendations near the document's opening. The presentation may be implicit rather than explicit, unclear, or disorganized.
- 1 – Writer makes little or no attempt to present major conclusions and/or recommendations near the document's opening.

Cohesion and Coherence

- 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.
- 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.

- 2 – Writing is fragmented: information flow between sentences and paragraphs is inconsistent. Sentences demonstrate little or no variety in style. Syntax may be irregular.
- 1 – Writing is incoherent and fragmented. Problems with syntax create barriers to reader understanding.

Grammar

- 4 – Writer establishes credibility with nearly perfect grammar, diction, and spelling.
- 3 – Memo contains some errors in grammar, diction, and spelling, but none that challenge reader understanding.
- 2 – Memo contains several errors in grammar, diction, and spelling that begin to hurt the writer’s credibility.
- 1 – Memo contains frequent or pervasive errors in grammar, diction, and spelling that create barriers to reader understanding and seriously hurt the writer’s credibility.

“Plain English”

- 4 – Writer uses active voice in nearly all of his or her sentences, and uses approximately 20 words per sentence (“Plain English” style).
- 3 – Writer uses active voice in most of his or her sentences, and uses a reasonable number of words per sentence.
- 2 – Writer uses passive voice frequently, and/or tends to wordiness.
- 1 – Writer uses passive voice predominantly, and/or writes excessively long sentences.

Reader Expectations

- 4 – Writer builds goodwill where appropriate by recognizing and meeting reader needs and expectations, using “you” language where appropriate.
- 3 – Writer usually builds goodwill where appropriate by recognizing and meeting reader needs and expectations, and usually uses “you” language where appropriate.
- 2 – Writer attempts to build goodwill, but sometimes misreads or fails to meet reader needs and expectations, or to use “you” language where appropriate.
- 1 – Writer makes no attempt to meet reader needs or expectations.

APPENDIX C.2**COMMERCE INTER-RATER RELIABILITY**

Table C.1. Inter-Rater Reliability

	Beginning of Course	End of Course
Plain English Style	0.36	0.50
Bottom Line Up Front (BLUF)	0.65	0.44
Clear Purpose	0.60	0.57
Cohesion & Coherence	0.05	0.00
Document Preview	0.65	0.72
Grammar	0.44	0.59
Reader Expectations	0.39	0.19
Skim Value	0.87	0.00

APPENDIX D

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 D.1):

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 D.2). 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%

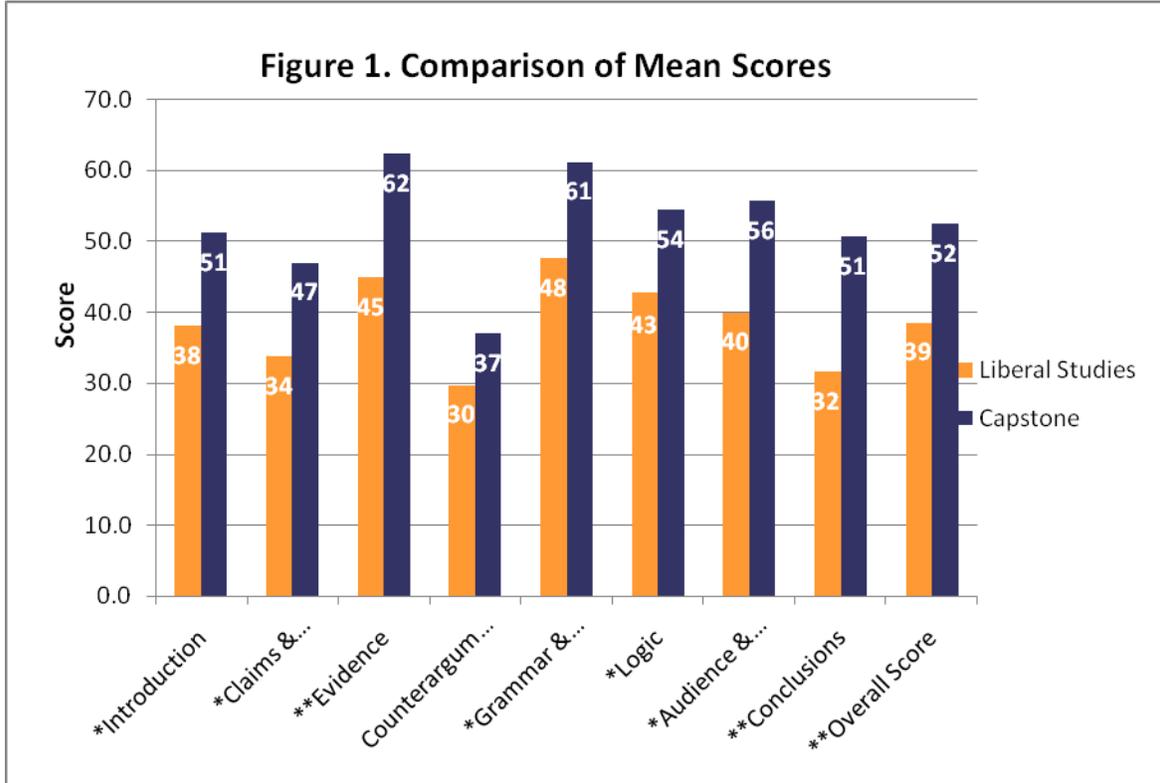
Table 3. Mean Scores on Learning Outcomes

	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 D.1

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 authorial tone. 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 D.2**BIS INTER-RATER RELIABILITY**

Table B.1 Inter-Rater Reliability

	Liberal Studies	Capstone
Introduction	0.32	0.50
Parts of Argument I – Claims & Subclaims*	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

