

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 were 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 were 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 A):

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). 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

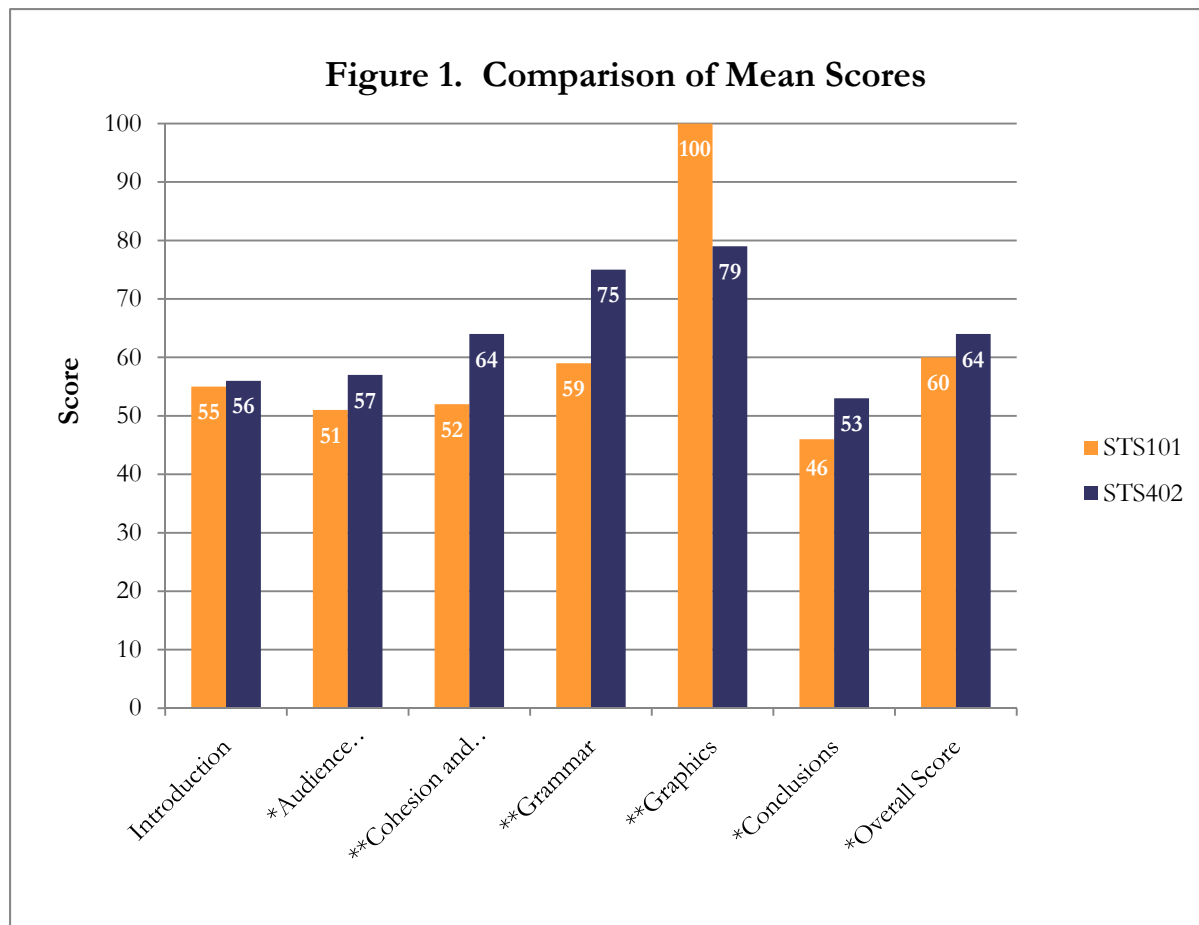
	STS 101	STS 402
Introduction	55	56
Audience Occasion and Purpose*	51	57
Cohesion and Coherence**	52	64
Grammar**	59	75
Graphics**	100	79
Conclusions*	46	53
Overall Score*	60	64

* 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 Cohe-
 nce learning outcomes. The smallest, but still significant, difference in performance was on
 Audience Occasion and Purpose (see Figure 1).



* 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.

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.

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

	Introduc- tion	Audience Occa- sion & Purpose	Cohesion & Cohe- rence	Grammar	Graphics	Conclu- sions
Introduction	.	.42**	.55**	.33**	-.22*	.46**
Audience Occasion & Purpose	.42**	.	.62**	.54**	-.09	.65**
Cohesion & Cohe- rence	.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.

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.

Table 6. Correlations among learning outcomes – STS 402

	Introduc- tion	Audience Occa- sion & Purpose	Cohesion & Cohe- rence	Grammar	Graphics	Conclu- sions
Introduction	.	.67**	.45**	.18	.28**	.54**
Audience Occasion & Purpose	.67**	.	.51**	.34**	.17	.58**
Cohesion & Cohe- rence	.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 $p < 0.01$.

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

Table 7. Correlations among learning outcomes - ABET

	Impacting global and societal con- text	Integration of contextual issues	Professional and ethical responsibil- ity	Role of engineering in contemporary issues
Impacting global and societal context	.	.67**	.43**	.70**
Integration of contextual issues	.67**	.	.49**	.68**
Professional and ethical re- sponsibility	.43**	.49**	.	.47**
Role of engineering in con- temporary issues	.70**	.68**	.47**	.

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

Appendix A

Six Learning Outcomes Developed by SEAS Faculty

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.

Four Learning Outcomes from ABET 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

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