

## **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 A):

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 B). 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

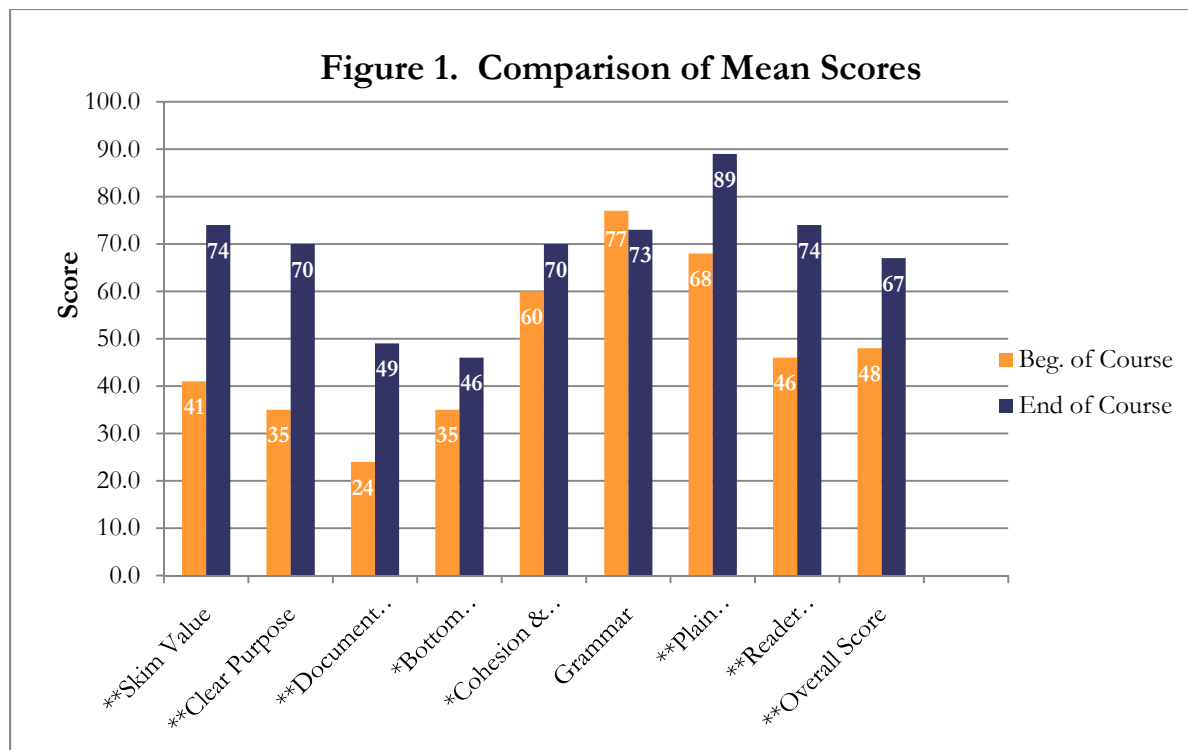
	Beginning of Course	End of Course
Skim Value**	41	74
Clear Purpose**	35	70
Document Preview**	24	49
Bottom Line Up Front (BLUF)*	35	46
Cohesion & Coherence*	60	70
Grammar	77	73
Plain English Style**	68	89
Reader Expectations**	46	74
<b>Overall Score**</b>	<b>48</b>	<b>67</b>

\*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 A

### McIntire Writing Assessment Rubric (Spring 09)

#### 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 B**

Table B.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