



UNIVERSITY of  
SAN FRANCISCO

**Educating Minds and Hearts to Change the World**



**Rhetoric and Composition  
Essay Assessment Results**

**Office of Institutional Assessment  
June, 2009**

## **Rhetoric and Composition Assessment of Student Learning Report, 2008-2009**

Mark Meritt  
Devon Holmes  
Leslie Dennen

### Introduction:

In the Fall of 2006, USF's Program in Rhetoric and Composition began an effort to assess annually the success of its freshman writing (RHET 110-120) and writing/speaking (RHET 130-131) sequences in helping students reach learning outcomes established for area A2, Written Communication, of the university's core curriculum. Though studies conducted for academic years 2006-2007 and 2007-2008 suggested modest gains in student learning over the course of the two semesters in both sequences, these positive assessments of our program were qualified by questions regarding the validity of our assessment methods (as described below). Though some questions regarding validity persist, the present study (for academic year 2008-2009), its investigators believe, has yielded more valid evidence of student progress toward two key learning outcomes. Also, the study's results raise questions for fruitful future inquiry regarding the factors that lead to greater student success in writing source-based prose.

### Method:

The method employed in this assessment project was in most ways similar to that used in the previous two studies. During the 2008-2009 academic year, the Program in Rhetoric and Composition collected first essays from fall semester sections of RC 110 and 130 (the first semester courses in the two sequences) and final essays from spring semester sections of RC 120 and 131 (the second semester courses in the two sequences) to be used as samples for our study of student progress toward our Core Area A2 learning outcomes. During the fall 2008 semester, two students from each section of RC 110 and 130 were randomly selected. Copies of these students' first essays (usually with grade and comments but without names) were to be collected by the Program in Rhetoric and Composition. At the end of the spring 2009 semester, these same students' final essays for RC 120 and 131 (with grades and comments but without student or instructor names) were also collected. Due to improved methods of and attention to gathering sample essays, we collected 30 pairs of 110-120 essays, and 21 pairs of 130-131 essays.

During the summer 2009 term, we began the process of reading and scoring the collected essays. Each essay was read twice, each time by a different reader. Also, readers recused themselves from reviewing work they recognized as produced by their own students. Readers assessed the degree to which each essay displayed evidence of the student's achieving two of the five area learning outcomes, rating each essay's performance in relation to each outcome on a scale of 0 (no evidence of mastery) to 3 (strong evidence of mastery). For each essay, the two readers' scores were combined to yield a final score on a scale of zero (0) to six (6). (See below for rubric and score result charts.)

Through this process, we hoped to determine the degree to which students' work and experience during our first-year writing and writing/speaking sequences enabled them to meet two of our core learning outcomes, specifically "integration of sources" and "style." More specifically, we hoped to see an overall improvement in scores (and therefore evidence of students' having met those learning outcomes) as students moved from the beginning of the first-semester course to the end of the second-semester course.

Due to questions regarding the validity and feasibility of the assessment method used for the two prior academic years, we modified our procedure in several ways. First, as noted above, we focused on assessing student progress toward only two outcomes (rather than four, as in previous years). We made

this decision for two main reasons. The simpler of the reasons was that it would make the process less time-consuming and allow us to focus more closely and meaningfully on a smaller area of interest rather than superficially address a broader field of concerns. The second reason was that our prior assessments of “critical analysis,” “integrating sources,” and “academic research” seemed based on invalid comparisons of essays addressing very different assignments that did not address the same outcomes or emphasize them to similar degrees. For example, early 110 and 130 assignments frequently required students to engage in critical analysis and perhaps to integrate one or two sources but rarely if ever required incorporation of multiple sources or even minimal research. Conversely, late 120 and 131 essays most often did require extensive research and integration of sources but rarely, if ever, required students to engage in close analysis of texts. These disparities in assignment task led to invalid comparisons between 110/130 and 120/131 essays, yielding invalid measures of progress toward learning outcomes or lack thereof.

As a result, we decided to assess student progress toward two outcomes: “integrating sources” and “style.” These two were chosen as likely to be addressed in both early and late assignments. In other words, any piece of student writing can be assessed in terms of style, and most of our sample essays integrated at least one text, whether assigned reading or research sources. Therefore, we could compare student performance in early and late essays according to similar criteria.

In order to further increase the validity of such comparisons, we revised our rubric for assessing students’ success in integrating sources. First, so that early essays (which typically integrated only one source [if any]) did not automatically receive lower scores based solely on their use of a very limited quantity of sources, we revised our criteria to exclude quantity of sources as a measure of student success in integrating source material, focusing instead on quality of integration of sources in the context of broader argument and analysis. Since most essays integrated one or more sources within the context of a larger writing purpose, the comparison of essays was more valid than in previous studies (though not entirely so, as will be discussed below).

Another factor that enhanced the validity of our study was our larger sample of essays. For the 2007-2008 study, we had obtained only a very small sample of 110-120 essay pairs, making it difficult to draw meaningful conclusions about student achievement in those courses. This year, we increased our sample size for 110-120 from four pairs to thirty.

Finally, we did not read paired essays (i. e. 110/120 and 130/131 essays by the same student) together. Instead, after assigning each essay a record number, we separated paired essays before reading them in order to avoid biasing scoring by building in an expectation of progress from earlier to later essays.

### Results, 110-120:

The average score on the “integrating sources” outcome for 110 students was 2. The average score on the same outcome for 120 students was 3.6, representing an increase of 1.6, or 80%. This increase is somewhat artificially inflated, as roughly one third of 110 essays did not require students to integrate sources at all. Such essays received a score of 0, drastically lowering the average 110 score on this outcome and thereby inflating the rate of progress between 110 and 120. Nevertheless, measurable progress for 110-120 students for this outcome remains apparent. If students whose 110 essay required no integration of sources are taken out of the sample, the remaining students averaged a score of 3 for this outcome in 110 and a score of 3.8 for this outcome in 120, an increase of 0.8 or 26%. Moreover, it should be noted that students whose 110 papers did not require integration of sources on average received a score of 3.2 for “integrating sources” on their 120 papers. This 0.6 (19%) positive difference in scores for students whose 110 assignments did require source integration may suggest that their experience writing

from sources from the beginning in 110 positively impacted their performance on this outcome in 120. However, the overall difference may not be significant. Also, several 110 students whose first papers did not incorporate sources went on to write 120 essays that scored quite well on this outcome. Finally, we do not know what kinds of follow-up assignments were given to 110 students whose first essays required no source integration. It is possible that among this small group (ten students) students whose second and third assignments required substantial engagement with sources scored higher than any students whose second or third 110 assignments did not require such engagement. In short, the discrepancy in scores between students whose first 110 assignment required source integration and those whose assignment did not suggests a need for further investigation both of a greater variety of assignments and of the possible relationship between differing assignment sequences and later performance in “integrating sources.”

The average score on the “style” outcome for 110 students was 3.6. The average score on the same outcome for 120 students was also 3.6. These identical scores at the beginning and end of the year may suggest a lack of progress in the category of style. However, since the writing demands of 120 essays (e. g., longer papers citing a greater quantity and variety of source materials often of a more challenging nature) are much greater than for early 110 essays, the maintenance of consistent scores across both terms may in fact suggest stylistic improvement, as composition scholars generally agree that stylistic and grammatical performance very frequently declines as students attempt more challenging writing tasks. In addition, when one includes scores only of students whose first 110 paper required incorporation of sources, the average score increases from 3.45 in 110 to 3.75 in 120. Conversely, when one includes scores only of students whose first 110 paper did not require incorporation of sources, the average score in fact decreases from 3.9 in 110 to 3.3 in 120. There are many ways to interpret this latter contrast. First, it seems to suggest that students write more fluently and without error when not incorporating sources in their writing. However, this conclusion is countered by the increase in style scores of students who incorporated sources for all papers. Also, since college writing (and most other forms of professional writing) requires integration of sources, leading students to produce comparatively error-free prose by not requiring source integration source is of limited value in an academic setting. Moreover, students whose 110 papers did not require integration of sources on average performed at a somewhat lower level (0.45) for the “style” outcome on 120 papers that required integration of sources, perhaps suggesting that their counterparts benefitted even on the level of style from more practice writing essays that required the integration of source material. Again, however, for reasons similar to those enumerated in the discussion of scores on “integrating sources” above, such an interpretation requires further study.

#### 110-120 Results, All Students

	110 score (average)	120 score (average)	Difference
Integrating Sources	2	3.6	1.6
Style	3.6	3.6	0

#### 110-120 Results, Students Writing Source-Based 110 Essays

	110 score (average)	120 score (average)	Difference
Integrating Sources	3	3.8	0.8
Style	3.45	3.75	0.3

#### 110-120 Results, Students Writing 110 Essays without Sources

	110 score (average)	120 score (average)	Difference
Integrating Sources	0	3.32	3.32

Style	3.9	3.3	-0.6
-------	-----	-----	------

### Results, 130-131:

The average score on the “integrating sources” outcome for 130 students was 3.09. The average score on the same outcome for 131 students was 4, representing an increase of 0.91, or 30%. These figures suggest that students began the 130-131 sequence with some skill in integrating sources, upon which they improved noticeably, if not dramatically.

The average score on the “style” outcome for 130 students was 3.71, while the average score on the same outcome for 120 students was 4.05, representing an increase of 0.34, or 9%. This gain in style appears modest. However, as noted above, maintaining a consistent score in stylistic fluency as students move on to increasingly challenging writing tasks may suggest stylistic improvement, as such tasks make greater demands upon the stylistic resources of student writers.

### 130-131 Results

	130 score (average)	131 score (average)	Difference
Integrating Sources	3.09	4	0.91
Style	3.71	4.05	0.34

### Comparing 110-120 and 130-131:

As might be expected based on their higher placement scores, 130-131 students as a whole performed better than their counterparts in 110-121 on both outcomes. The average 120 score for “integrating sources” was 3.6, while the average score for 131 students was 4. However, it is noteworthy that 110-120 students whose 110 papers required integration of sources averaged a 120 score of 3.8, 50% closer to the 131 average. All 130 papers required integration of sources. These facts may suggest that 110-120 students (with some exceptions) who work more with sources earlier in their first-year composition sequence perform at a level closer to that of 130-131 students. Again, however, the issue requires further investigation for several reasons: the sample of 110 students writing papers citing no sources was very small, some of those students scored well on source integration in 120, and examination of intervening assignments was not part of this study.

The average 120 score for “style” was 3.6, while the average score for 131 students was 4.05. Again, it may be worth noting that the average on this outcome for 120 students whose 110 papers required source integration was 3.75, while those who whose 110 papers did not require integration of sources was 3.3. In fact, as noted above, the average “style” score for 120 students whose 110 papers did not require source material actually decreased from 3.9 to 3.3, perhaps suggesting that some of these students’ sentence-level fluency “buckled” under the pressure of having to write from sources. Again, however, such an interpretation requires further inquiry.

### Comparing 08-09 with 07-08:

The performance of 110-120 students on the two outcomes assessed was stronger this year than in the 2007-2008 academic year. In 07-08, 120 essays earned an average score of 3 on “integrating sources”

and a score of 3 on “style.” In contrast, 120 essays this year scored an average of 3.6 in both “integrating sources” and “style,” representing a 20% increase in performance. This improvement may be the result of improved instruction or higher admissions standards. However, the 110-120 sample from 2007-2008 was very small, making general conclusions difficult.

In 2008-2009, 130-131 students performed more strongly in “integrating sources,” averaging a score of 4 for 131 essays, as opposed to a 3.4 average for 2007-2008. In the area of “style,” 130-131 students performed at a slightly lower level in 2008-2009 than in the previous year, the score for 131 essays dropping from 4.4 in 2007-2008 to 4.05 in 2008-2009.

### Discussion:

Overall, this assessment project produced reasonably good news about student progress toward the two learning outcomes on which we focused our attention (integrating sources and style). Students’ scores on these outcomes improved significantly, if not dramatically for both groups (110-120 and 130-131). Though 110-120 students’ scores on style remained steady over the course of the year, this consistency, as noted above, suggests that students’ writing style is not showing expected or typical signs of decline as they face more challenging tasks (a development one might reasonably assess as “improvement”). Scores also indicate that 110-120 students whose first 110 papers required integration of sources scored better on average on both outcomes in 120 than did students whose 110 papers did not require source integration, though it is not clear whether the margin of difference is sufficiently significant to draw general conclusions. This difference in scores (between those of 110-120 students whose early work integrated sources and those of students whose work did not) points to a potentially rich field of investigation for future assessments.

Though we found our assessment overall this year to be more focused and more valid (due to changes introduced), we still saw several possibilities for improvement. To a great extent and in different ways, particularly in comparing 110 and 120 essays, we still seemed to be comparing very different types of writing, making the validity of the project still somewhat questionable. For example, comparing personal narrative papers from some 110 students to 120 research-based essays was not useful for assessing progress in the skill area of integrating sources. Also, even when comparing source-based 110 essays with source-based essays in 120, the comparison still did not seem entirely valid, as the 110 papers were frequently close reading or rhetorical analysis assignments, while the 120 essays were frequently research-based arguments. These two types of genres both “use” or integrate texts, but they do so in very different ways and for different purposes. A more valid study would compare analytical papers from the two courses with each other, or research-based arguments from the two courses with each other.

In addition, even though we refined and improved our rubric for assessing integration of sources, it became apparent that we could further do so. For example, our guidelines tended to favor essays that treated sources more analytically or critically and to underrate essays that used sources more straightforwardly as evidence to support an argument. Also, essays that quoted or paraphrased excessively could score reasonably well as long as they provided some minimal commentary. The rubric could perhaps be refined to grant more credit to more judicious and meaningful citation of source material.

We can significantly improve our assessment of “style.” Currently our rubric calls for rather vague and impressionistic assessments of sentence structure. To a great extent, style and grammar may be assessed more rigorously and quantitatively, perhaps through methods like measurements of error frequency and of sentence length and complexity.

Finally, in reading and scoring the essays, we agreed that, if possible, we should obtain clean copies of student essays that are free of teachers' grades, editorial marks, and commentary. These marks and comments were at times distracting or confusing, and could perhaps have biased the scoring process. Obtaining clean copies should be relatively easy, as teachers can simply photocopy requested essays before marking them or may even print out clean copies submitted electronically (or forward them electronically themselves).

### Numerical Results Charts:

#### 110-120: Integrating Sources

Student number	110 score (combined)	120 score (combined)	Difference
01	2	2	0
02	2	3	1
03	4	3	-1
04	4	4	0
05	0	5	5
06	0	2	2
07	3	6	3
08	0	3	3
09	4	4	0
10	2	6	4
11	2	4	2
12	3	2	-1
13	0	3	3
14	2	2	0
15	4	2	-2
16	2	2	0
17	5	5	0
18	3	5	2
19	5	4	-1
20	0	4	4
21	0	5	5
22	0	2	2
23	0	2	2
24	2	5	5
25	2	4	2
26	1	5	4
27	0	4	4
28	0	2	2
29	4	3	-1
30	4	5	1
Average	2	3.6	1.6

## 110-120: Style

Student number	110 score (combined)	120 score (combined)	Difference
01	2	2	0
02	3	3	0
03	4	4	0
04	3	4	1
05	4	4	0
06	1	2	1
07	4	5	1
08	4	3	-1
09	3	3	0
10	4	6	2
11	3	3	0
12	4	2	-2
13	4	4	0
14	3	4	1
15	4	4	0
16	3	4	1
17	4	5	1
18	2	4	2
19	5	4	-1
20	5	4	-1
21	4	3	-1
22	5	3	-2
23	4	3	-1
24	3	4	1
25	4	4	0
26	4	4	0
27	4	4	0
28	4	3	-1
29	4	2	-2
30	3	4	1
Average	3.6	3.6	0



130-131: Integrating Sources

Student number	130 score (combined)	131 score (combined)	Difference
01	2	4	2
02	2	3	1
03	3	3	0
04	4	4	0
05	3	5	2
06	3	5	2
07	3	3	0
08	4	6	2
09	4	4	0
10	4	5	1
11	2	4	2
12	4	4	0
13	5	3	-2
14	2	5	3
15	5	*	*
16	5	3	-2
17	2	6	4
18	2	3	1
19	2	4	2
20	2	2	0
21	2	4	2
Average	3.09	4	0.91

## 130-131: Style

Student number	130 score (combined)	131 score (combined)	Difference
01	4	5	1
02	3	3	0
03	3	4	1
04	4	4	0
05	5	5	0
06	4	5	1
07	4	2	-2
08	4	4	0
09	4	4	0
10	4	5	-1
11	4	4	0
12	4	4	0
13	4	4	0
14	4	4	0
15	4	*	*
16	4	4	0
17	3	4	1
18	4	4	0
19	2	4	2
20	4	4	0
21	2	4	2
Average	3.71	4.05	0.34