

## **W Course Assessment at the University of Connecticut, 2008**

A Project of the General Education Oversight Committee

Report composed by Tom Deans

### W Assessment Participants

*Art History:* Duncan Givans, Faculty Coordinator; Robin Greeley, Associate Professor; Michael Orwicz, Associate Professor

*English:* Tom Deans, Associate Professor; Kathleen Tonry, Assistant Professor; Mandy Suhr-Sytsma, Graduate Assistant

*Human Development and Family Studies:* Lisa Kraimer-Rickaby, Faculty Coordinator; Louisa Kimball, Graduate Assistant; Samit Bordoloi, Graduate Assistant

*Political Science:* Virginia Hettinger, Faculty Coordinator; Daniella Melo, Graduate Assistant; Chris Paskewich, Graduate Assistant

Statistical analysis by Scott Brown, Professor of Educational Psychology, and Donalyn Maneggia, Program Assistant, Teachers for a New Era

### **Summary**

This report focuses on direct assessment of the writing done in advanced W (writing-intensive) courses, which are an integral part of UConn's General Education requirements. During summer 2008, a team of faculty and graduate students completed an outcomes-based assessment in cooperation with the Art History, Political Science, and Human Development and Family Studies departments. During winter 2009, we will complete the Nursing assessment and will issue a follow-up report.

The good news: We discovered that in W courses UConn students are composing long, source-driven papers on intellectually challenging topics. We determined—drawing on the expertise of nine faculty and graduate students scoring anonymous papers in their home disciplines—that of the 128 final papers collected for this study, 94% met at least minimal expectations for advanced writing in the major. That so few papers were rated unsatisfactory is likely a consequence of policies that keep W course size small, allowing ample student-faculty interaction, and that require revision, prompting faculty get involved early in each student's writing process.

The not-so-good news: The overall quality score for 83% of papers fell between “minimally proficient” and “moderately proficient,” which means that we saw a large clustering in the low-middle range. When comparing student performance by year in college, we did not find evidence that seniors are writing better papers than sophomores or juniors. We also noted that instructor grades did not correlate significantly with rubric scores, and that instructor grades were

significantly higher than independent reader scores. All this suggests the need to set the bar higher in our W courses and our grading; we should also explore how to align the writing-intensive components of the UConn curriculum.

The rubric scoring revealed the relative strengths and weaknesses of student writing in Art History, HDFS and Political Science courses. Those findings are being cycled back to the departments for discussion and action, and we have already seen some improvements in W course design and delivery as a consequence of this study, although we hope to see more as the results are more widely discussed.

The faculty and graduate students involved endorsed extending this assessment process to other academic departments not only because the study delivered findings useful to shaping more effective W courses but also because it allowed participants sustained opportunities to reflect on the UConn curriculum and on their own teaching. The faculty development embedded in this assessment is one of the project's immediate outcomes.

This report describes the research design, summarizes and interprets results, and makes recommendations for instructors, departments, GEOC, and UConn administrators. An appendix with supporting materials is also included.

## **I. Guiding Assumptions and Research Design**

Several assumptions guided the design of this assessment project:

- We would use the *department* as the unit of analysis, not individual instructors, course sections, or students.
- We would do *outcomes-based* assessment—that is, evaluate what students, in general, could do as academic writers by the end of a W course. We did not try to measure growth across the span of a W course or to inventory the content of W courses (though such assessments could be valuable).
- We would focus mainly on *direct assessment of student writing* rather than on indirect measures, such as course syllabi or surveys.
- At the same time, we would use *diverse methods* to collect and evaluate data. These included collecting final course papers, collecting the instructor grades for those papers, having students fill out a self-efficacy questionnaire, scoring papers with rubrics, running a statistical analysis of all that data, doing deep qualitative evaluations of selected papers, and reserving time for extended discussions within and across departmental clusters.
- The process would be *led by faculty, driven by dialogue, and open to revision*.
- We would be *future-oriented*—aimed at sparking evidence-driven discussions about teaching, learning and curriculum design in the participating departments. Our evaluation would be more formative than summative.
- The process was *attentive to the complex nature of writing*. That is, we approached writing less as a set of discrete skills that lend themselves to atomized testing than as a complex, context-sensitive mode of learning, communicating, and doing.

- We understood academic writing as *bound to disciplinary context* and therefore we customized our evaluation tools, especially the rubrics, to reflect departmental values and priorities.
- We wanted this study to be as much about *faculty development* as about assessment of students. The study presented a rare opportunity for faculty from different fields to sit together and do sustained inquiry focused on student writing. Indeed, the insights and the relationships that emerged from two weeks of intensive collaboration in June were an immediate and durable outcome of the project.

Preparations began in the fall of 2007 with conversations in the Assessment Subcommittee of the GEOC, and by spring a research plan was approved by the Institutional Review Board at UConn. The original plan was to include four academic departments, but that was reduced to three because Nursing offers W courses only in the fall—therefore we delayed involving Nursing until fall 2008.

Our plan hinged on collecting three types of data: final student papers in W courses, instructor grades, and data from a student questionnaire. We intended to analyze it in five ways: scoring with discipline-specific rubrics; performing statistical analysis of relationships among the rubric scores, questionnaire items, and instructor grades; completing a formal error inventory; doing intensive source-checking and qualitative analysis of a subset of papers; and engaging in sustained qualitative discussions within and across the departmental clusters.

Here is the plan as mapped out in the IRB form:

This study aims to assess the learning and writing performance of students enrolled in a set of writing-intensive (W) UConn courses. As part of a larger and ongoing effort by the Faculty Senate's General Education Oversight Committee to assess general education, this study is designed to help us better understand how students are learning and writing in W courses, to allow comparison of actual student writing to university-wide expectations and department-specific expectations for W courses, and ultimately to be of use to participating UConn academic departments as they reflect on and improve their W courses. The research will involve the voluntary collection of samples of student writing from W courses; it will also include a student questionnaire that invites participants to self-report on their own practices of and attitudes toward writing.

This is not a pre/post study. It is intended to provide a snapshot of the nature and quality of student writing being done in W courses. The findings from this study may be used as a baseline for future W course assessments at UConn.

We will assess student writing using rubrics that reflect university and department learning outcomes for W courses; we will furthermore ask students to complete a questionnaire on self-perceptions of writing competencies to see how they correlate with the quality of their writing as scored on rubrics.

Procedures:

1. Do informed consent.
  - a. Informed consent form explained in class by study representative; course instructor will not be present
  - b. Note: student identification is confidential after collected papers, questionnaires, and instructor comments have been linked using a code. Once the items have been coded, the names will be redacted and the

informed consent forms will be detached from the data and stored separately.

2. Obtain end-of-semester writing sample from each student – we will attempt to collect all electronically in a web-based drop-box, but in some circumstances we may need to collect some paper writing samples. Samples will be collected before instructors comment on them.
3. Administer the writing self-efficacy instrument to all students in the classes. The informed consent and survey will take place during the first 15-20 minutes of a class. Students who opt not to participate will be invited to take that time to prepare for the day's class.
  - a. Administer during the last week of class
    - i. record names of students to link to those papers being analyzed
  - b. In the future, administer the scale twice
    - i. during the first week of classes
    - ii. during the last week of classes
4. Develop scoring rubrics
  - a. One tied to university W goals and criteria
  - b. One tied to discipline specific issues. The four faculty coordinators from each of the four participating departments will create a rubric that speaks to the concerns of their own discipline and their own department's W courses.
5. Train raters of papers (raters will be graduate students from the 4 participating departments, trained and supervised by Deans; faculty coordinators from the 4 departments will be available to handle discipline-specific training and rating concerns)
  - a. Teach inter-rater agreement
6. Trained raters score papers with rubrics, blind to instructor grades and comments; papers will be identified by codes rather than names.
  - a. Will be done after the semester and well after course grades are submitted.
7. Analyses
  - a. Tally rubric scores and determine patterns in within each discipline and across all four
  - b. Gather and record qualitative observations of the readers/raters to account for any student writing strengths and weaknesses, if any, that are not captured by the rubric categories
  - c. Compare rubric scores to scores of instructor
    - i. correlation
  - d. Correlate the self-efficacy measures with the rubric scores
    - i. correlations
  - e. Group/discipline differences
    - i. ANOVA for the disciplines
      1. rubrics
      2. self-efficacy
  - g. If time allows, we will also randomly select 4-8 student papers from each of the 4 disciplinary groups of papers and do an audit of how the writers use their sources in those papers. This will test the claims of a recently published article that students tend to draw most of the quotations they use in papers from the first third of the articles or books they use; it will also assess whether they are using their sources in sound academic ways.

In early spring 2008, we recruited faculty coordinators from Art History, HDFS, and Political Science. They worked with Tom Deans, their home departments, and each other to create rubrics that reflected the priorities for advanced undergraduate writing in their respective disciplines (for the three rubrics, see Appendix A).

At the end of the 2008 spring semester we visited twelve W sections (3 AH, 3 HDFS, 6 PS) and administered the informed consent and self-efficacy questionnaire. The overall yield of papers was 128, with 30 from AH, 31 from HDFS, and 67 from PS.

We did not collect any papers from W sections at the regional campuses. The Freshman English writing assessment planned for spring 2009 will collect writing from all six campuses, and that precedent should be followed for future W assessment initiatives.

Roughly 2/3 of students in the twelve sections consented to participate. We cannot know for sure how such self-selection influenced our results, though it no doubt did. We might speculate, for example, that some students who knowingly plagiarized opted out, and this should be kept in mind when academic honesty is discussed later. We did not attempt to collect any senior honors theses, which are counted as W courses, and those would likely have pulled up the quality of the sample.

We gathered our team of twelve faculty and graduate students in June 2008. We started by discussing a set of articles on writing assessment by Kathleen Blake Yancey, Brian Huot, Sandra Murphy, Edward White, and John Bean (see Appendix B for readings and the June schedule). This allowed us to arrive at shared understandings of how such terms as “validity” and “reliability” are used in assessment, to survey best practices in the field, and to question the assumptions and goals of our own plan.

The following activities were also completed during those weeks in June:

- Did practice readings and scoring sessions to arrive at inter-rater reliability.
- Scored papers in clusters by discipline, using the rubrics. Each paper was scored independently by at least two readers.
- Analyzed papers within departmental clusters.
- Read papers across the three departmental clusters and discussed as a full group.
- Did deep audits (comprehensive source checks) of a subset of 23 papers.
- Attempted a formal error inventory, without success.
- Reflected on our overall findings and recommendations.
- Reflected on the assessment process and possible next steps.

Statistical analysis of the rubric scores, questionnaire items, and instructor grades was done by Scott Brown, Professor of Educational Psychology, and Donalyn Maneggia, Program Assistant, Teachers for a New Era (see Appendix C for their report).

The fourth department in this project, Nursing, delayed participation because all Nursing W courses are offered in the fall semester. As of November 10, 2008, sixty-five nursing students from five Nursing W sections have completed the questionnaire and have consented to allowing their writing and grades to be collected at the end of the semester. Scoring and analysis are scheduled for January 2009.

## II. Results and Discussion

1. W Course Instruction and Staffing
2. Overall Student Writing Outcomes By Rubric Scores
3. Overall Student Outcomes From Qualitative Analysis
4. Department-Specific Writing Outcomes
5. Instructor Grades in Relation to Departmental Expectations for Writing in the Major
6. Grammar, Style, and Documentation  
*Formal Error Analysis*
7. Deep Audit Results  
*How Students Are Using Sources*  
*Academic Honesty*
8. Student Self-Efficacy Results
9. Writing Development Across Four Years at UConn
10. Dissemination of Results

### 1. W Course Instruction and Staffing

Because the study focused on student learning outcomes, we did not evaluate W course teaching or content. Still, we collected the syllabi of the twelve sections from which we solicited papers and noted that all met the letter and spirit of University expectations for W courses. All required at least 15 pages of writing across the span of the semester and all featured structured opportunities for revision.

Nine of the twelve courses required a long, research-driven final analytical paper or literature review, assigned in stages, with instructor feedback provided on at least one draft. Page length for the end-of-semester assignments ranged from 4 to 27 pages; the average length was 14 pages. The average number of sources incorporated was 9. Overall, the paper topics struck our faculty and graduate assistant readers as relevant and intellectually challenging, even when the arguments or reviews of literature were not well executed.

One concern is that instructional staffing for these courses was often at odds with General Education policies. General Education guidelines state: “W Courses will normally be taught by University of Connecticut faculty. When that is not possible, then qualified students may be used to assist faculty in 200-level courses or, with faculty supervision, teach a 100-level W course.” Of the twelve courses involved in this study, six were taught by tenure-track or tenured faculty, three were taught by assistant professors in residence, and three were taught by graduate students leading sections linked to a common lecture taught by a long-term adjunct instructor.

### 2. Overall Student Writing Outcomes By Rubric Scores

Based on direct assessment of 128 final papers, and using the holistic rubric score for writing in a particular major as our measure, 94% of UConn students met at least minimal proficiency for writing in their major.

The mean score for holistic quality was 2.51, midway between “minimal proficiency/2” and “moderate proficiency/3” on a 4-point scale. Few papers (8 of 128) received an overall “unsatisfactory” score; few (14) received an overall “excellent” score. That means that there was significant clustering in the low-middle: 83% of students submitted final papers judged as minimally or moderately proficient academic writing in the major.

That only 11% of papers were rated as excellent discipline-specific writing may be due in part to the very high standard that scorers set for that category, but this is still clearly an area UConn should target for improvement.

That so few students (6%) wrote unsatisfactory papers is encouraging and was likely influenced by UConn's W policies: enrollment is capped at 19, allowing for ample faculty-student interaction; and a revising process is required, ensuring that professors can intervene early in the writing process. Research on writing pedagogy consistently shows that especially this latter practice—formative feedback provided during the writing process—is pivotal in helping novice writers grow and succeed.

### **3. Overall Student Outcomes From Qualitative Analysis**

Student learning and writing are complex phenomena. Assessing them with rubrics, even very good ones, usually will not tell the whole story. Therefore, we built time into our study to ask, "What patterns did we see across the full sample that were not captured by the rubrics?" and to document our general observations. The following themes that emerged from our cross-disciplinary discussions:

- We were generally impressed with the ambition of UConn instructors and students in taking up challenging topics for research and writing.
- Grammar did not—to the surprise of some—prove an obstacle to our comprehension or scoring of the papers.
- We noted patterns of genre mixing. For example, a fair percentage of political science papers adopted features of journalistic discourse, which sometimes complemented and sometimes conflicted with expectations for academic analysis and argument, and we observed some HDFS students taking the approach of a hands-off survey of research while others gravitated toward a thesis-driven approach akin to that of an article or critical essay (the HDFS literature synthesis assignment called for both practices). Many students seemed to be mixing genres more accidentally than self-consciously, and this suggested to us less the need to privilege one genre over another—unless that is what the department or instructors intend—but instead a need to make genre itself a point of discussion, analysis and emphasis in W courses.
- Introductions were fairly reliable indicators of how the paper would unfold. Writers of overly general or muddled introductions tended not to recover; writers of focused and relevant openings tended to follow through of the promise of their introductions. This was less about whether or not a student included a thesis (most did) than about how the writer situated the thesis in the introduction.
- Where we saw an economical, lively, and eloquent prose style, we usually saw a strong argument or analysis. That is, style and substance, more often than not, went hand in hand. This was affirmed by a statistical analysis of rubric scores revealing that the 'style/language' rubric item was the best predictor of the rubric holistic quality score.
- We saw an under-appreciation for how defining terms (in sophisticated ways) can set up and advance academic analysis. All too often writers failed to define key terms for readers or they did not use them consistently as they developed an argument.
- Another frequent problem was writers failing to signal clearly the shifts between their own voices and those of others. This was less a documentation issue (students

generally cited their sources) and more a lack of awareness about how to use transitions and meta-discourse to orchestrate a complex analysis—to signal clearly when they are advancing their own arguments and when they are summarizing or riding on the arguments of their sources.

- Even when they used good, peer-reviewed academic sources, students were generally too timid in weighing their sources critically, putting sources in conversation with one another, and inserting themselves into the scholarly conversation.
- The session during which we read papers from outside our home disciplines affirmed for most that we should, in future assessments, continue doing rubric scoring in departmental groups. Participants valued reading papers from other disciplines but affirmed the value of rubric scoring being done by insiders in a given field.

Video of the June 2008 qualitative discussions is available. If interested in seeing it, please contact the University Writing Center.

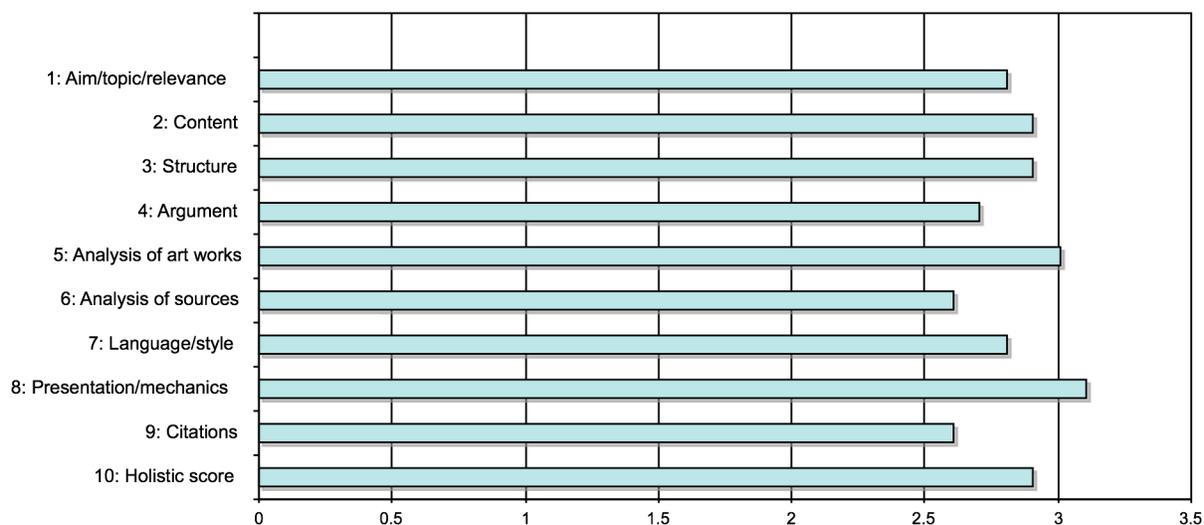
#### **4. Department-Specific Writing Outcomes**

In large universities, departments are the place where courses and curricula are shaped. In this study we emphasized a process that would cycle what we learned back to departments. This process started in spring 2008 as each faculty coordinator invited colleagues in his or her home department to deliberate on what to include in a discipline-specific rubric. It continued during two intensive weeks in June as seven faculty members and five graduate students scored papers and engaged in sustained dialogue. All of us have since returned to our home departments with valuable experience and knowledge to share.

The tables and graphs that follow reveal the comparative strengths and weaknesses of student writing in each of the three departments. All scores are a result of independent rubric scoring by faculty and graduate students from those same departments. The papers were read without instructor comments on them or any knowledge of the grade; readers also had no access to the assignment. For a full articulation of what the 1/2/3/4 rating scale means for *each* departmental rubric and *each item* on a rubric, please see the expanded rubrics in Appendix A. The rubric scores should prove immediately useful to departments as they assess their own majors and discuss what to emphasize in their W courses.

We should resist comparing rubric means across the three departments because we practiced for inter-rater reliability *within* each departmental cluster of scorers, not across the three departmental clusters. Lower or higher overall scores may indicate more lenient or more rigorous scoring within a given departmental cluster. We should focus instead on the ten rubric items *in comparison to each other* (as relative strengths and weaknesses within one department) and *in comparison to departmental standards for writing* (as detailed in the expanded rubrics in Appendix A).

<b>ART HISTORY RUBRIC</b>	<b>Means (based on 30 papers)</b>
<b>1. AIM OF PAPER:</b> Identifies and addresses topic directly; topic addressed corresponds to assignment; topic, claim or question addressed is relevant for its readership and appropriate in scope	2.8
<b>2. CONTENT OF PAPER:</b> Content of paper justifies argument/position adopted; identifies and addresses relevant range of critical/scholarly perspectives; identifies and addresses sufficient range of pertinent artistic material	2.9
<b>3. STRUCTURE OF PAPER:</b> Uses sections and paragraphs to give logical structure and fluency to development of argument; uses introductory paragraph(s) to set out question/thesis to be examined; uses body of text examine question/thesis; uses concluding paragraph(s) to resolve question/thesis	2.9
<b>4. ARGUMENT OF PAPER:</b> Argument clearly expressed and sustained throughout paper at the level expected; argument based on appropriate balance between analysis of art works and analysis of scholarship; use of quotation to strengthen argument, rather than simply cull factual information; demonstrates awareness of pertinent issues outside immediate art historical context of paper	2.7
<b>5. ANALYSIS OF ART WORKS:</b> Argument demonstrates critical analysis of artworks at level expected; argument demonstrates sensitivity to potential and limitations of artworks under consideration, both in themselves and as discrete categories of evidence	3.0
<b>6. ANALYSIS OF TEXTUAL SOURCES:</b> Argument demonstrates critical analysis of relevant literature at level expected; makes appropriate distinctions between classes of evidence and source, e.g. artist's statements, contemporary critics and later commentators; primary and secondary sources etc	2.6
<b>7. LANGUAGE:</b> Evinces a prose style, a tone, word/terminology/language choices, verb tenses, syntax, and other stylistic moves appropriate to academic writing in art history at the undergraduate level	2.8
<b>8. PRESENTATION:</b> Grammar, mechanics, spelling, punctuation, proofreading, and formatting	3.1
<b>9. CITATIONS:</b> Accurate and consistent use of references, citations, and bibliography in keeping with departmental guidelines (Chicago Style)	2.6
<b>10. HOLISTIC RATING:</b> Assessment of the paper as work of art history, both in its broadest sense and in the particular form engaged by the topic and genre	2.9



1=Unsatisfactory    2=Minimal Proficiency    3=Moderate Proficiency    4=Excellent/Outstanding

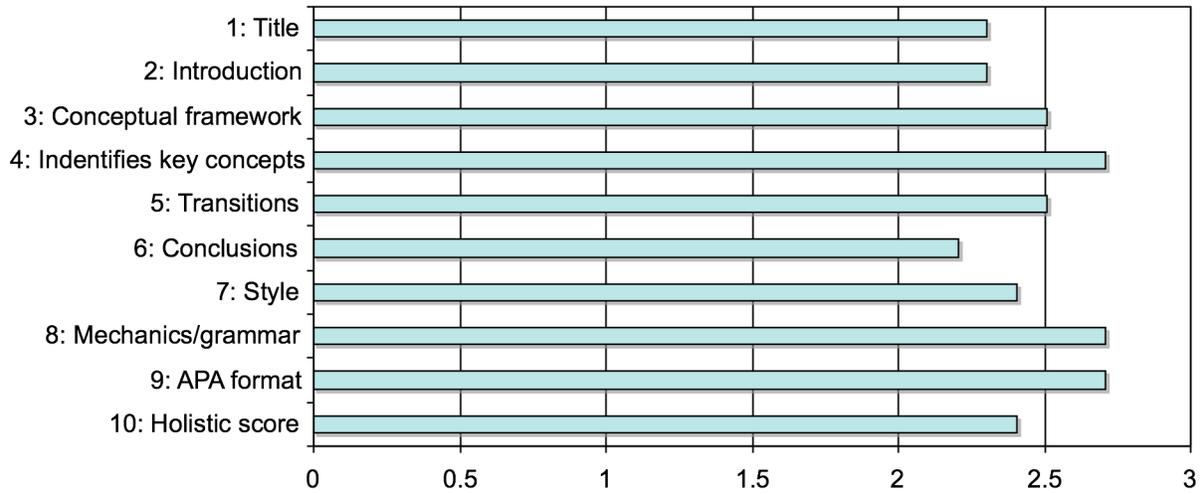
For Art History all rubric item means fell between *minimally proficient* and *moderately proficient*, though closer to *moderately proficient*. That suggests no consistently outstanding features and no glaring problems in the student writing. Students were strongest in analyzing works of art and in editing for presentation/grammar; students were weakest in using secondary/scholarly literature in the field and in citing sources. The AH readers also saw a need for improvement in formulating a thesis, defining terms, and organizing an extended analysis. They were pleasantly surprised by the high degree of student engagement with the material, and readers from PS and HDFS were generally impressed by the sustained analysis they saw in the Art History papers.

\*

\*

\*

<b>HDFS RUBRIC</b>	<b>Means (based on 31 papers)</b>
1. Title conveys subject and focus of the synthesis	2.3
2. Introductory paragraph adequately introduces the topic and thesis of the synthesis	2.3
3. Synthesis goes beyond a simple summary of the articles cited to develop a conceptual framework that links summaries and articles	2.5
4. The synthesis clearly identifies a set of key concepts	2.7
5. Transitions tie sections together, as well as adjacent paragraphs	2.5
6. Conclusion connects main topic to overall points and gives a general consensus	2.2
7. Evinces a prose style, a tone, word/terminology/language choices, verb tenses, syntax, and other stylistic moves appropriate to academic writing in HDFS at the undergraduate level (note especially apt use of active voice and use of past tense in references to past research)	2.4
8. Grammar, punctuation, and spelling	2.7
9. Adheres to APA format in text and on reference page	2.7
10. Holistic score: Assessment of the paper as a work of HDFS, both in its broadest sense and in the particular form engaged by the topic and genre	2.4



1=Unsatisfactory 2=Minimal Proficiency 3=Moderate Proficiency 4=Excellent/Outstanding

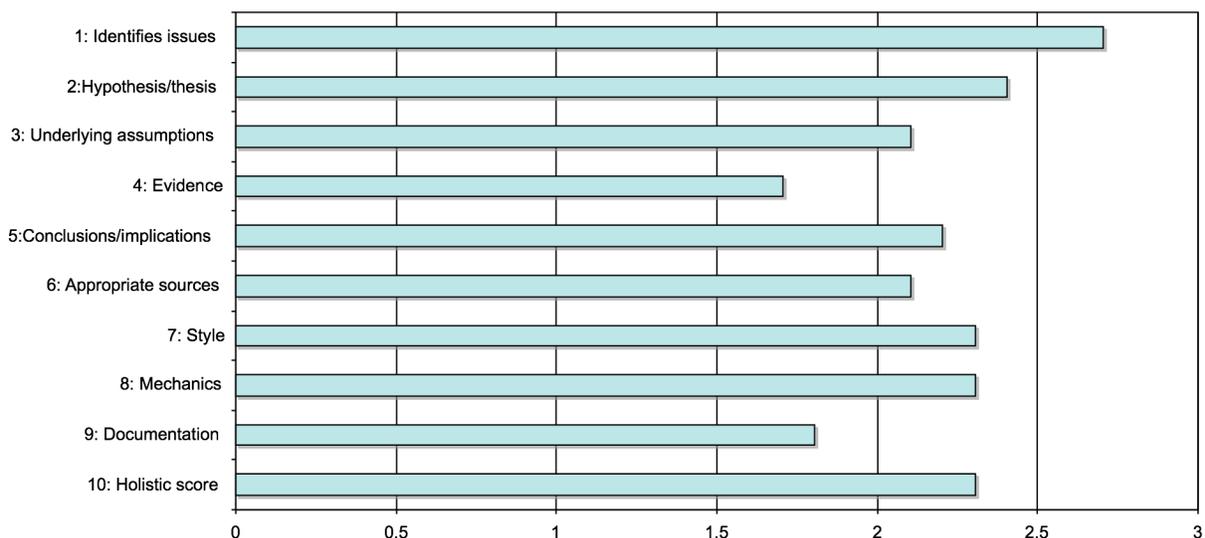
For Human Development and Family Studies all rubric item means fell between *minimally proficient* and *moderately proficient*. That suggests no consistently outstanding features and no glaring problems in the student writing. Students were strongest in identifying a set of key concepts and sub-concepts, as well as in editing for readability, tone and flow. The lowest scoring rubric mean was the one keyed to how students developed a theoretical framework and arrived at conclusions: “Conclusion connects main topic to overall points and gives a general consensus.” While all students surveyed a series of topically related scholarly articles, not enough of them introduced an assertive thesis, distinguished clearly between their own ideas and those of their sources, or discussed the implications of the thesis in the conclusion. This was consistent with what readers from PS and AH observed, although they were impressed with the quality of the sources HDFS students selected (almost exclusively peer-reviewed journal articles) and with how students documented those sources (responding to a demand that they use APA style, HDFS students followed documentation conventions much more closely than students in the other two departments).

\*

\*

\*

<b>POLITICAL SCIENCE RUBRIC</b>	<b>Means (based on 67 papers)</b>
1. Identifies the main issue(s) in a given (existing) argument	2.7
2. Articulates one's own significant hypothesis/opinion/thesis clearly, using fitting experience and/or external sources for argument	2.4
3. Identifies underlying assumptions and alternative perspectives of the argument	2.1
4. Assesses the quality of the supporting evidence on the issue(s)	1.7
5. Identifies conclusion/implications of the issue/argument beyond the example given	2.2
6. Uses appropriate sources in support of one's argument	2.1
7. Evinces a prose style, a tone, word/terminology/language choices, verb tenses, syntax, and other stylistic moves appropriate to academic writing in political science at the undergraduate level	2.3
8. Mechanics are correct, including grammar, syntax, spelling, and punctuation	2.3
9. Adheres to academic/political science standards for the appropriate attribution of others' work	1.8
10. Holistic score: Assessment of the paper as work of political science, both in its broadest sense and in the particular form engaged by the topic and genre	2.3



1=Unsatisfactory 2=Minimal Proficiency 3=Moderate Proficiency 4=Excellent/Outstanding

For Political Science, two rubric item means fell below *minimally proficient* and the rest between *minimally proficient* and *moderately proficient*. Students were strongest in identifying the main issues in their argument and in articulating a thesis; they were weakest in documenting their sources and assessing the quality of the supporting evidence. The PS readers also saw a general need for more critical thinking in the papers. Readers from AH and HDFS valued how the political science students were more comfortable than their own students in making assertive arguments.

While AH and HDFS papers were drawn from three W sections each, PS papers were drawn from six different W sections, and the instructor assignments and expectations for those sections varied more widely than in the other departments: three sections required 4-7 page reflective or journalistic analyses that cited few or no sources; three sections required 15-25 page papers that incorporated 10+ sources. The rubric was designed with the longer, source-driven critical analyses in mind; therefore, the reflective and journalistic papers often earned unsatisfactory ratings on the three rubric items keyed to evidence, pulling down those means. Inconsistency in student documentation/citation was likewise, in part, a reflection of political science as a diverse collection of subfields with different approaches to research and documentation. Yet there were some documentation patterns—such as students not understanding when and how to paraphrase or document such paraphrasing—that went beyond varied citation conventions in the field. Given this range of course expectations, it would have been helpful to have the PS assignments.

### **5. Instructor Grades in Relation to Departmental Expectations for Writing in the Major**

Instructor grades for papers averaged much higher than the holistic/summary rubric score. The mean instructor grade for the full sample was 3.61/4.0; the mean rater holistic score was 2.51/4.0. In plain terms, many students are getting A and B grades for work we judged as less than moderately proficient.

This may be simple grade inflation, yet several other interpretations are also possible. Our scorers read the papers stripped of context—without even the assignment sheets—and much grading is influenced by those assignments and by the instructor's individual priorities. Moreover, grading is often used, at least in part, as a motivational tool, with credit given for showing effort, following course protocols, showing progress across drafts, displaying growth across the semester, and so on. Institutional factors are also likely in play: there are few institutional incentives for rigorous grading. In fact, quite the opposite: in most departments student teaching evaluations are the only measure of teaching merit, and lower grades can bring lower teaching evaluation scores. Teaching evaluations are especially consequential for untenured and contingent faculty, who taught half of the W sections from which we collected papers.

Just as jarring as the grades/rubric score gap is that the correlation between overall rubric scores and instructor grades was .134 and was not statistically significant. In other words, across our full sample, instructor grades and rubric scores on the same papers showed no statistically significant relationship one another. This problem was more acute in Art History and Political Science because the topics and course requirements for those sections were more diverse. Also, no instructors in those departments were using rubrics like the ones we used in our scoring (because those rubrics did not exist before late spring 2008). While grade inflation in HDFS was slightly more pronounced than in the others, the correlation between grades and scores was tighter and statistically significant. That better correlation may be likely a consequence of the

three HDFS sections being taught by the same instructor (albeit with three TAs who did grading), using a common final assignment, requiring a common genre (the literature synthesis), and graded with a common rubric (one very close to the HDFS rubric used for this study).

Another curious grading pattern is that those taking a second W course received higher grades than those taking a first W course even though paper quality was not rated any higher on the discipline-specific rubric. Similarly, seniors received higher grades on their papers than underclassmen, even though paper quality was not rated higher. Perhaps rubrics are not capturing something (academic content that is not keyed to writing?). Perhaps seniors are more skilled than underclassmen at “doing school” or at addressing all the factors (other than paper quality) that typically go into a grade.

Clearly these two issues of high grades and inconsistent correlation between grades and what the departments say they value (as reflected in the rubrics) merit attention. Within departments, grading practices should be made more transparent and discussed more openly.

As a direct consequence of the spring/summer assessment project, some encouraging progress on aligning grading practices with departmental writing expectations is already evident:

- In the Political Science department, the rubric developed for this assessment project is being used in some W courses to make expectations clear to students and to guide grading. Use of the rubric is being encouraged but not mandated, and the number of sections using it is still in the minority. Still, discussions of this assessment study may encourage more to adopt it, or some aspects of it, in the near future.
- In a key HDFS W course with multiple sections, the instructor has adopted a practice learned during the summer of 2008: having graduate assistants for the W sections score practice papers, discuss the scores, and repeat the process until arriving at sufficiently calibrated inter-rater reliability. This has led to grades that are not just lower but also more consistent across sections and better aligned with departmental expectations for undergraduate writing.

## **6. Grammar, Style, and Documentation**

Students see grammar as a weakness: in an open-ended question on the self-efficacy survey, they listed grammar more frequently than anything else. And several faculty and graduate scorers entered this project expecting to have their reading consistently frustrated by tangled grammar and prose. However, in our cross-departmental qualitative discussions toward the end of the study, the group concluded unanimously that the frequency and seriousness of grammar, syntax and documentation errors did not, in general, obscure intent or meaning of the student writing. Readers observed many irksome errors in grammar and style, as well as many departures from conventional academic documentation, but they agreed that those problems seldom were serious enough to impede comprehending the writer’s line of development.

Rubric scores suggested a similar story about sentence-level editing. Art History, HDFS and Political Science rubrics shared three common criteria: grammar/mechanics/presentation, style/language/toner, and citations/source documentation. Across the full sample, the average score for grammar/mechanics/presentation was 2.6, which falls midway between *minimally proficient* and *moderately proficient*.

As for documentation, when a consistent style was stressed in both instruction and grading—as was the case with HDFS stressing APA—students rose to that expectation consistently. In

contrast, for Political Science, a field in which several sub-disciplines demand different documentation conventions, we saw both syllabi and students giving less attention to citation conventions (though student writers usually gave readers some sort of a path back to the original course) and perhaps as a result, documentation was the lowest mean rubric score (1.8). The Political Science department has since used this finding as cause to set a shared documentation standard for its W courses.

We defined *style* not only in general terms (control, economy, and variety of language; transitions/“flow”; clarity) but also as the degree to which students adopted the register of the discipline for which they were writing. The average score for style was 2.5. Of all the rubric items, style was the best predictor of the holistic score. This may mean that those students who are already sophisticated in their use of language are those most inclined to do well overall in academic writing; it could also suggest that instructors should attend more to style in their teaching as one way to leverage a higher percentage of excellent papers.

#### *Formal Error Analysis*

Our assessment plan included doing a formal error analysis, which is an inventory of the kinds and frequency of grammar and usage errors. We modeled our approach on a study published in 1988 by Robert Connors and Andrea Lunsford. Unfortunately, we were unable to complete the formal error analysis because we exhausted the time set aside for it in our (unsuccessful) efforts to arrive at inter-rater reliability with practice papers. Four faculty and five graduate students, representing four disciplines (AH, HDFS, PS, and English), simply could not arrive at a reasonably consistent consensus about what should be labeled as an error. This surprised us all, as we assumed that this process would be relatively easy.

This failure precipitated a robust discussion about how what seems like a simple practice—marking papers for grammar errors—is in fact much more complicated and vexing than most assume. This experience suggested the need for more faculty development on this issue.

This does not mean that we should altogether abandon an error analysis. Certainly counting surface errors constitutes only a small part of the intellectually generous definition of writing we assumed for this study, but it would still be interesting to see how UConn students compare to college students nationally on this specific measure.

Since the summer of 2008, a new benchmark study on error analysis has been published. Tom Deans will use its method to analyze the writing we have collected, plus the writing to be collected from Nursing in fall 2008 and from Freshman English in spring 2009. That combined formal error analysis is scheduled for summer 2009.

## **7. Deep Audit Results**

We conducted “deep audits” of 23 student papers. This involved faculty and graduate assistants doing sustained qualitative reviews of papers in their home disciplines. Because this process included tracking down and reviewing all the sources cited in a given paper, each deep audit took 2-3 hours.

### *How Students Are Using Sources*

When examined under that kind of microscope, nearly half of the papers (11 of 23) were rated overall as “poor” in their use of source material. We note that the nature of this process led the reviewers to be especially rigorous in their evaluations. Still, the deep audits pointed to several areas we need to improve when teaching research writing.

First, some encouraging data:

- The average number of sources used in these 23 papers was ten.
- The average number of peer reviewed journal articles included per paper was seven. Students used peer review journal articles more frequently (by far) than any other kind of source.
- Student comprehension of sources (6 poor | 17 satisfactory or excellent) was rated as slightly better than how they put those sources to use (8 poor | 15 satisfactory or excellent).

Consistent with other studies that track the research habits of undergraduates, UConn students are almost exclusively using online databases and electronic versions of journals rather than browsing the stacks or using hard copy bound journals.

We also evaluated *why* and *how* students were using sources by taking an inventory of the *purposes* to which students put sources to work in their papers:

- To supply background/contextual information (21 of 23 “sometimes” or “often” did this).
- To support a thesis (21 of 23 “sometimes” or “often” did this).
- To serve as template for the writer’s own argument (16 of 23). This is not meant to denote plagiarism but instead the habit of novice writers to track a bit too closely to one of the sources they admire (as opposed to staking out their own innovative or original claim).
- No clear reason (12 of 23).
- To introduce dissenting points of view (10 of 23; this was also highest “never” tally).

We introduced two of those categories—“to serve as template for the writer’s own argument” and “no clear reason”—only after doing some practice readings and letting the patterns emerge.

That more than half of our students incorporate sources that serve no purpose in the paper—other than, we presume, to meet the requirement to include X number of sources—and that more than half of them never use a source to introduce a dissenting point of view suggest at least two clear points that should be addressed when instructors teach research writing. UConn students seem to have gotten the message that they need to use peer reviewed sources in their papers. And their comprehension of those sources seems reasonably good. Now instructors need build on those promising habits by explaining and modeling how real scholarly conversations and arguments work, and by insisting that each source should serve a clear purpose.

### *Academic Integrity*

While doing the deep audits we discovered breeches of academic integrity, and that is always distressing, but our findings suggest that students in UConn W courses are doing

at least as well or better than national averages on measures of academic honesty. We cannot declare that with great confidence, however, since our sample was small and 1/3 of those invited to participate in this study did not consent (and we can reasonably assume that those students bent on cheating would not have consented).

Among the 23 papers for which we did deep audits, we found four that included acts of gross plagiarism—that is, cases we judged as intentional and fraudulent. We found ten more papers in which at least one a source was misused, used questionably, or paraphrased by not cited—but we interpreted these ten cases less as acts of fraud than as lapses of diligence or consequences of inexperience, confusion, or ignorance (of academic conventions). We found no cases of a student inventing a source or buying a paper from an online paper mill.

Clearly we need to do more in our W courses to teach students to use and document sources, and perhaps we need to do more to catch cases of gross plagiarism. Yet some of the news is encouraging, and the UConn policies capping W courses at 19 and mandating that faculty get involved in each student’s writing process are both good for academic integrity.

## **8. Student Self-Efficacy Results**

Students who consented to participate completed a self-efficacy questionnaire that asked them to rate their writing abilities and share aspects of their own writing process (see Appendix A). The data gleaned from that questionnaire, when analyzed in relation to their rubric scores and grades, suggested that students’ own perceptions of their writing abilities are not significantly at odds with rater and teacher perceptions of their writing.

When invited to list their own strengths and weaknesses on open-ended questions, the most frequent student responses for weaknesses were grammar, research, introductions/conclusions; the most frequent responses for strengths were clarity, research, and evidence/support.

As we add more students to our data set, we will be able to say more about self-efficacy and especially about student writing process issues. For example, do students who visit faculty in office hours or go to the Writing Center get higher rubric scores or grades? Our sample of 128 was too small to arrive at conclusions about such questions, but we may be able to answer them as we increase the sample size by adding more departments to this assessment project.

## **9. Writing Development Across Four Years at UConn**

Longitudinal studies of writers in college typically follow a cohort of students over several years, and the few that have been done suggest some cautions for making claims about writing development over several years: (a) writing development is rarely steady and linear, even for students in carefully aligned vertical curricula (and we should note that the planning, linking, teaching and assessing of UConn’s writing requirements—Freshman English, plus two W courses—have never been formulated or coordinated as a vertical sequence); and (b) that causality (course X resulted in Y writing skills) is vexingly difficult to discern because writing development is entwined with overall intellectual development, personal growth, student motivation, and student socio-economic circumstances. Our study took one snapshot of one moment in time, thus limiting even further what we can claim.

With those cautions in mind, we can approach what our data told us about writing performance across years at UConn:

- The overall quality rubric scores of those taking a second W course were not higher than those of students taking a first W course (although their grades were).
- The rubric scores for grammar/mechanics, documentation/citation, and style (the only criteria consistent across all three rubrics) for those taking a second W course were not higher than those of students taking a first W course.
- The overall quality rubric scores by seniors were not higher than those of undergrads (although their instructor grades were).
- Those who took Freshman English at UConn did not score better on their W papers than those who took Freshman English elsewhere.

Such results are disheartening and suggest a need for greater alignment both between Freshman English and W courses and among W courses within and across departments. Planning for such alignment should also take into account the realities how students typically move through their UConn writing sequence, especially the typical pattern of incoming students taking Freshman English in the first year but then, because of limited W course seat availability, not having access to W courses until their junior or senior year.

When we look to learning outcomes for students finishing a W course (the main focus of this study), UConn students seem to be doing pretty well. Common sense would suggest that Freshman English and W courses contribute to those good outcomes, yet this study offers no evidence to confirm if or how writing-intensive components of UConn's General Education build on one another in a cumulative way.

The assessment of student writing in Freshman English, scheduled for spring and summer 2009, will allow us to read the writing done in Freshman English alongside that done in W courses. While still not as good as a multi-year longitudinal study following the same students, a comparison of the Freshman English assessment and the W assessment should give us better evidence make claims about writing development across years at UConn.

## **10. Dissemination of Results**

Findings from this study have been disseminated to audiences both within and beyond the UConn community, although more dissemination is still needed and anticipated.

The University Writing Center hosted a presentation of the results for the university community on October 9, 2008 and invited all faculty (20 attended). On November 22, 2008, Tom Deans, Lisa Kraimer-Rickaby, and Louisa Kimball presented a paper based on this study at the 2008 Quinnipiac Biennial Conference on Writing and Critical Thinking. Both presentations were well received. Moreover, Tom Deans will be working on ways to publish results in composition studies and writing across the curriculum journals.

Disseminating results to Art History, Political Science, and HDFS faculty as a way to spark evidence-driven discussions about departmental W courses remains a priority. The faculty coordinators had planned to present at faculty meetings in each department in the early fall of 2008, but the budget crisis at UConn so dominated faculty concerns that we thought it wise to delay. Still, Virginia Hettinger of Political Science presented an overview of the Political Science findings to her colleagues; a Political Science brown bag event to present more results and discuss them is being scheduled for February. Meetings in Art History and HDFS are being

planned for early in the spring semester. Another audience that may be interested in the results is UConn librarians. The report will be shared with them too.

Some outcomes of the study are already evident in the participating departments. The rubrics, for example, are being now used by some W instructors (we don't know exactly how many) to announce departmental expectations for writing and guide grading.

### **III. Recommendations**

Recommendations for the Art History, HDFS, and Political Science departments should be clear in the “Department-Specific Writing Outcomes” section above: faculty in the departments should address how to maintain the strengths in student writing confirmed by this assessment and discuss how to change pedagogy and curricula to address the systemic weakness. What follows are general recommendations for departments, instructors, GEOC, and the University administration.

#### **Recommendations for Academic Departments**

1. Introduce some form of direct assessment of student writing that reaches across W sections. This study confirmed the value of this approach not just for program assessment but also for faculty development. Direct assessment of writing need not involve doing a formal study such as this one. It could be done on a smaller, less formal, less time-consuming scale by collecting twenty or so papers from the department's W courses and having a committee read or score them and report back to the department. Such practices could also prove useful as part of the departmental assessments required by the Provost or by outside accrediting agencies. The University Writing Center is available to consult with departments on conducting such assessments.
2. Work toward some degree of alignment across W offerings. Departments can do this through discussion, sharing assignments, and/or crafting a document or rubric that spells out department-specific priorities for writing in the major. The rubrics created by Art History, HDFS, and Political Science can serve as models. Of course, departments cannot require faculty to use a common rubric but they should at least encourage a shared awareness of department-wide expectations for writing. A departmental document or rubric on writing would be valuable for future assessments and for addressing the distressing gap between grades and rubric scores discovered in this study.
3. Have open and honest—even if difficult—discussions in the department about grading in W courses.
4. Consider introducing a session of practice paper scoring—especially when multiple sections of the same W course are being offered and TAs are doing the grading—to arrive at a reasonable degree of inter-reader reliability. One HDFS course with multiple sections has recently implemented this practice (after learning it during this assessment process) and has seen significant improvements in grading consistency.

#### **Recommendations for Instructors of W Courses**

1. Continue doing what is working: assigning long, source-driven papers that engage students in challenging and relevant intellectual work, and supervising an active revision process that is helping nearly all students meet at least minimal proficiency.

2. Focus more on the criteria in the top halves of the AH, HDFS and PS rubrics (purpose, argument, analysis, evidence, structure, conceptual framework) and style (language and register of the discipline) rather than on grammar or mechanics. Sentence-level correctness is important, and expectations for it should be high, but we did not find grammar and mechanics the most pronounced problems of UConn student writers.
3. Give genre expectations more attention. As they travel through various courses, students inevitably get conflicting messages about expectations for academic writing because different disciplines have different conventions and different ways of arguing. Since W courses are introducing students to writing in a particular discipline, they should make disciplinary and genre expectations explicit. That may mean devoting some class time to shepherding students through samples published academic writing, with the focus on the *rhetorical moves* of the text—on *how* experienced academic writers introduce claims, include counter-arguments, employ transitions, use meta-discourse to guide readers, and bring an argument to a close. Introductions and conclusions should get special attention, as many papers went astray there.
4. Emphasize and model how defining key terms is often vital to setting up academic analysis and advancing an argument.
5. Work toward helping students not just find, select and comprehend sources (which they are doing reasonably well) but *assess* those sources critically and *use* them purposefully in a sustained analysis. Deep audits suggested that too many students use sources uncritically as templates for their own claims and not enough of them use sources to introduce dissenting points of view. Bringing reference librarians into dialogue about these issues would be wise.
6. Use student writing samples in teaching—not only as models but also as prompts for discussion and analysis. This study makes samples of student writing in PS, AH and HDFS available, and selected papers of varying quality will be converted into PDFs for use in teaching. Instructors should also collect samples from their own students for use in future teaching.

### **Recommendations for GEOC and Writing Assessment Study Faculty Coordinators**

1. Complete the original plan for the study by including a fourth department (Nursing) in the study. 65 Nursing students across five W sections have consented to participate and have completed self-efficacy questionnaires; papers will be collected by mid-December 2008; and scoring is scheduled for January 2009. An updated version of this report that includes the Nursing data will be issued in spring.
2. Follow through on the original vision to use study results to provoke discussion about teaching and curriculum, particularly in the four participating departments.
3. Complete the formal error analysis.
4. Compare the findings from this study with the findings from the writing assessment being done in Freshman English in spring and summer 2009. This is a necessary prelude to discussing alignment between Freshman English and W courses.
5. Support expanding this project to include additional UConn departments and the regional campuses. The faculty and graduate instructors who participated in this study valued it not only for what we learned about the quality of student writing but also for the opportunities it opened for cross-disciplinary, data-driven dialogue about teaching and learning. All reported that the experience has enriched their own teaching and approach

to curriculum design. In short, the project served as both program assessment and faculty development.

6. Resist the impulse to create a UConn-wide writing rubric for W courses or assessment. Our experience confirmed what much of the research in writing across the disciplines tells us: that while academic writing may share some general features, it varies quite significantly by discipline, and faculty are best positioned to assess writing within their home disciplines.
7. Return to Art History, HDFFS, Political Science and Nursing in three or four years to repeat our assessment, using results from this study as a baseline.
8. Continue building an archive of student writing. This will be useful for research and for teaching (i.e., convert selected student papers from the archive to PDFs and make them available to W instructors for use in teaching).

### **Recommendations for UConn administrators**

1. Provide continued funding through GEOC for the reader/scorer stipends so that additional rounds of this assessment process can be undertaken with more academic departments. If those stipends were provided, the University Writing Center could handle administering the assessments.
2. Explore ways that academic departments might better align W courses with each other. If that proves too unwieldy, offer incentives for departments to create a vertical sequence of two W courses within their own major (a few departments already do this).
3. Find institutional ways to reward rigorous grading. Expanding this assessment process to other departments is one way to do this; another may be to wean UConn off student teaching evaluations as the main tool for evaluating instructors.
4. Maintain support for the 19 enrollment cap for W sections to ensure that faculty can address writing in process.
5. Share this study with outside accrediting agencies that value direct and outcomes-based assessments of student learning that relies on more than grades.

### **Appendices**

Appendix A: Rubrics, consent form, self-efficacy questionnaire

Appendix B: June schedule, readings

Appendix C: Statistical analysis

Appendix D: Deep audit form and samples