

## Summary Report on the Assessment of Academic Writing at the University of Connecticut 2010

During the past two years two general education assessment projects—one focusing on W courses and another on Freshman English—have helped us to better understand how UConn students are performing as writers. Those efforts resulted in four separate reports: *W Assessment Report, 2008*; *NURS W Assessment Report, 2009*; *Report on Writing Assessment in Freshman English, 2009*; and *Report on Writing Assessment in Electrical and Mechanical Engineering, 2010* (all are available online at [geoc.uconn.edu/assessment](http://geoc.uconn.edu/assessment)). This report is essentially a meta-analysis of those earlier reports, synthesizing and summarizing findings that pertain to writing across the UConn curriculum. It also features a fresh statistical analysis of the aggregate data gathered from the six participating W departments, resulting in amendments to several findings originally reported in the 2008 W report, which was based on data from only three departments.

These studies were *outcomes based assessments*—that is, they attempted to measure, based on a careful reading and scoring of final course papers, what students could do, as writers, by the end of a Freshman English or W course. The studies were not designed to measure how students arrived at their writing competency or what they learned during a particular semester, although the Freshman English assessment was able to say something about that because early and late papers were collected. Nor was the research designed to measure progress across 4 years at UConn, although the W study collected some data that allowed us to comment in a limited way on longitudinal development.

We can discern several patterns when we place the Freshman English and W outcomes assessments alongside one another, especially because the studies used a common four-point scale to rate papers both holistically and on various sub-skills of academic writing. Yet the two assessments cannot be easily grafted to deliver a single measure of how students are writing in their first year as compared to later years at UConn because the Freshman English project assessed how students were writing *in relation to the goals of Freshman English as determined by English faculty and graduate student readers*; likewise the W assessment for Political Science assessed how students were writing *in relation to the standards sets by the Political Science department as measured by Political Science faculty and graduate students*; and so on with the other five participating departments (Art History, Human Development and Family Studies, Mechanical Engineering, Electrical Engineering). We cannot measure writing in absolute terms because writing is a deeply contextual activity. No single, stable, specific definition for “good writing” holds consistent across the disciplines, something we discovered in the course of these assessments and that is affirmed by the consensus of published research on academic writing.

A full explanation of the methods for the W and Freshman English (FE) studies can be found in earlier reports. Note that the methods of the W and FE projects were similar but not identical. For example, for the W project instructor grades were collected, which was not part of FE study; for the FE project instructor assignments were collected and scored, which was not part of the W studies. Similar philosophies guided both assessments: we aimed to do *outcomes-based* assessment—that is, evaluate what students, in general, could do as academic writers by the end of a given course or major; we focused on *direct assessment of student writing*; we used *diverse methods*, both quantitative and qualitative; while attentive to best practices for validity and reliability, the project was *driven by dialogue* among faculty and by context-sensitive measures rather than by decontextualized tests; our approach was *more formative than summative*, aiming to spark evidence-driven discussions about teaching, learning and curriculum in the participating departments; and we tried to be *attentive to the complex nature of writing*—that is, we approached writing less as a set of atomized skills and more as a context-dependent mode of learning and communicating.

This report excludes much of what was most useful in the four earlier reports, especially the department-specific findings about student writing that informed specific recommendations for teaching, course design, and curriculum. Still, when putting the earlier assessment reports alongside one another to discover which outcomes are most generalizable for the University of Connecticut, several key patterns and conclusions emerge.

### **Measures of Proficiency**

Proficiency for writing in a given discipline was defined by faculty in that department and codified in a 10-item rubric. 4 criteria were common across all the rubrics (style, grammar and mechanics, citation practices, and a holistic score); 6 criteria were customized to reflect each department's priorities and therefore differed across departments. Each paper was rated independently by at least two faculty or graduate students scorers who used a scale of *unsatisfactory*, *minimally proficient*, *moderately proficient*, and *excellent*.

- *Based on the blind review holistic scores, the vast majority of students—84% for Freshman English and 93% for W courses—are submitting at least minimally proficient writing that is appropriate to the course and discipline.*
- *For both Freshman English and Ws, most final papers scored between minimally proficient and moderately proficient on most rubric items, including the holistic score.*
- *Faculty and graduate student scorers set a high bar for minimal proficiency, indeed higher than most UConn instructors. This was affirmed by an analysis of instructor paper grades in relation to blind rubric scores: nearly all papers scored “unsatisfactory” by the assessment project received passing instructor grades. Scorers likewise set a high bar for the moderately proficient and excellent categories. This is affirmed by an analysis of instructor grades on papers collected in relation to blind rubric scores: only a small minority of W papers given an A grade by instructors earned an excellent score from the blind reviewers. While grades were not collected for the Freshman English assessment, the scoring was similarly rigorous; only impressive intellectual work, expressed in well-crafted prose, earned high scores.*

- *The low percentage of papers scored unsatisfactory (16% in Freshman English, 7% in W courses) suggests that Freshman English and W courses are working well in helping students of varied preparedness make gains and find success in academic writing.* That few students are falling through the cracks may be due to reasonable section enrollment caps (19 for W courses, 20 for Freshman English), W and FE policies that require formative feedback on drafts, dedicated instructors, and sound teaching practices. Research on academic writing suggests that such factors improve learning, allowing instructors to catch struggling writers early and mentor them toward proficiency.
- *For both cohorts, few final papers received a blind holistic rating of excellent: 4% in FE and 9% in Ws.* Ideally, the percentage of work rising to the *excellent* category should be higher. The greatest potential for moving students from minimal and moderate proficiency toward excellence would be helping them master higher-order concerns while still attending to editing, style, and documentation (which could use improvement, albeit to a lesser degree).

### **Performance on Writing Sub-Skills**

If we trust faculty and advanced graduate students in a given discipline to be the best judges of undergraduate writing in their fields, the holistic score is the most valuable measure of overall writing competency. Yet we also collected data on how students are doing on various sub-skills in writing. Three of those sub-skills (style, grammar/mechanics, citation practices) were measured across all departments; six varied by department. The following findings proved most persistent across the seven participating departments.

- *Higher-order concerns, such as doing analysis, building an argument, applying theory, weighing evidence, synthesizing sources, and drawing conclusions stood out as the biggest shortfalls in the papers, making them the logical points of emphasis for course design and faculty development.* This was affirmed by both rubric scoring and qualitative discussions. These areas, which blur traditional boundaries between “writing,” “critical thinking” and “content,” are at the heart of both developing writing competence in a given discipline and achieving a broad liberal education.
- *Grammar and mechanics are not the most pressing writing problems for UConn students.* Some scorers were surprised to find that sentence-level errors rarely impeded their ability to read, comprehend, and evaluate the student writing, which in nearly all cases had been through a revising process. Grammar was not even close to the biggest problem. FE students scored higher on *grammar/mechanics/correctness* than on any of the other seven rubric items (*inquiry, defined project, textual engagement, rhetorical knowledge, organization and development, style/voice, holistic score*). For W papers the overall grammar/mechanics mean rubric score was 2.7 on a 4-point scale; the median was 3.0. This suggests room for improvement, but for no department was grammar/mechanics the lowest scoring sub-skill.
- *Each department’s relative strengths and weaknesses are summarized in the earlier assessment reports.* Because 6 of the 10 criteria for each scoring rubric were department-specific, the data available to each individual department is richer and more reliable than the comparative data across departments. This study was premised on the assumption that

academic writing at UConn is best assessed at the departmental level, by faculty. While adapting rubrics to what each department values in student writing may frustrate quantitative, cross-university analyses, our primary aim was to deliver information that would help departments understand and improve the teaching of writing in their specific disciplinary contexts. Extensive qualitative discussions with the faculty and graduate student participants revealed some insights not captured by the rubrics; those varied by department and are detailed in earlier reports.

### **Writing Development Across Years at UConn**

As noted earlier, the studies were not designed to capture rich data on how students develop as writers over the years of the college careers. However, statistical analysis of the W data suggests mixed results on this.

- *UConn seniors vs. Underclassmen.* As measured by *blind holistic rubric scores*, there is no statistically significant difference between the performance of seniors and underclassmen on their final W papers. As measured by *instructor grades*, however, UConn seniors are performing better than underclassmen.
- *Students Taking a First W Course vs. Students Taking a Second W Course.* There is no statistically significant difference in the *blind holistic rubric scores* of students who took a W course earlier and those who did not, except on one rubric item: students taking their second W course do slightly better with grammar and sentence-level editing than those taking their first W course. Students taking a second W course received higher *instructor paper grades* than those taking their first W. These seemingly contradictory findings might be explained by the fact that W courses are not part of a vertical sequence—that is, they are not designed to build upon one another; they teach different disciplinary conventions, different styles, and different citations practices. The apparent lack of overall writing development is disappointing, but given this context it makes some sense that transferable skills from one W to the next would be limited to sentence-level editing and a generalized sense of how to meet teacher expectations for grading.
- *There were no statistically significant differences in the rubric scores of students who completed English 1010 or 1011 at UConn compared to those who did not, although those taking 1010 elsewhere rated themselves higher on a self-efficacy questionnaire.* That means that those students taking first-year English elsewhere thought that they were better at most writing tasks even though their actual W paper performance was not. This suggests that when compared to other places where UConn students get Freshman English credit (ECE courses in the high schools, AP credit, transfer credit from other colleges), the UConn's Freshman English program seems to better tamp down overconfidence and give students a more realistic sense of their academic writing abilities. Also notable is that incoming students with ECE and AP credits tend to come from more socioeconomically privileged backgrounds, which suggests that Freshman English may be serving an equalizing function, helping to bring all UConn students, including a disproportionate number from less advantaged backgrounds, into accord with high expectations for college writing.
- *In both Freshman English and W courses, students are composing substantial papers in*

*response to challenging assignments.* Final papers for FE averaged 7 pages and engaged with difficult readings; final papers for W courses averaged 13 pages and employed multiple outside sources. The following passage from the Freshman English assessment report seems to hold true as well for W courses: “We can report that the required Freshman English courses are in the main vigorous courses with substantive reading and writing components and an attention to writing as a process of engagement, reflection, and revision.”

### **Patterns in Grading**

While grading was not a focus of these studies, we did record the grades that W papers received and compared them to the rubric scores. “Instructor grades” here is used to signal the grade the paper received—whether from a tenured faculty member, graduate assistant, or adjunct instructor—and not the course grade. The following observations are based on comparing instructor grades to blind holistic scores.

- *Instructor grades for W papers were higher than blind scorer ratings of the same papers.* The instructor grades for final W papers were typically about a full letter grade higher than the independent score grade equivalents, suggesting a need for more rigorous grading in W courses.
- *Instructor grades for W papers correlated with blind scorer ratings for those same papers.* The correlation for all six W departments was .293 with a p-value of less than .05. This shows that instructor grades reflect the departmental writing values (as expressed in the rubrics) reasonably well, although not as strongly as we might hope. The correlation was stronger in departments that expected one consistent genre, like the literature review or lab report, to be assigned across W sections; it broke down for departments that had little consistency in the genres assigned by W instructors, little communication among those instructors, and no common writing rubric. An important qualification is that most departments were creating their rubrics during the same semester when we were collecting papers, which means that W instructors had no access to the rubrics. In other words, they weren’t using the rubrics to set course expectations or grade papers. If departments engage in open discussion of their assessment findings and begin using the rubrics developed as part of this project in their W sections, their correlations should grow stronger.

### **Academic Integrity and Patterns in Source Use**

The W assessments involved extensive source checking for a subset of each department’s papers (10-20% of the total); in cases like Electrical Engineering where students were composing lab reports based on experiments rather than research papers based on sources, we reviewed the integrity of student data collection and use.

- *Academic integrity is encouraging.* Source and data checking revealed very few instances of gross plagiarism, no paper mill papers, and no made-up sources; there were, however, many instances of source misuse that crossed the line into plagiarism but that scorers attributed to a lack of understanding or care rather than to intentional fraud. This suggests that W courses need not extra policing but instead a greater emphasis on more sophisticated strategies for source use and documentation. As with other findings of this

study, sample bias probably affected these results: students had to consent to participate in this study, and students who intended to plagiarize would have been unlikely to consent. Still, the findings on academic integrity are encouraging.

- *The deep audits revealed that students are doing generally acceptable work in finding, evaluating, and using sources.* On the upside, they are moderately proficient at finding reliable sources, using them to establish background, and deploying them to support their claims; on the downside, they rarely use outside sources to introduce counter-arguments and often include some sources for no apparent reason other than to meet the requirement for a prescribed number of sources. Citation formatting was just minimally proficient (mean for all Ws was 2.1, the median 2.0), but this varied by department: in some departments, such as HDFs, students faithfully followed citation conventions because it was part of the assignment expectations and instruction; in departments that did not articulate expectations or offer guidance, students cobbled together non-standard systems.

### **Putting Writing Assessment to Work for Faculty and Students**

The assessment studies summarized here were intended to be *formative* than *summative*; were meant to propel evidence-driven discussions of teaching and learning—department by department—rather than deliver summary judgments about UConn students or the UConn curriculum.

- *Outcomes assessment can provoke teaching changes.* Participating departments have used the findings from these projects to improve their teaching practices. The changes have been incremental but important. For example, the art historians discovered that their students were weaker than they had hoped in using secondary sources and therefore built an additional library module into their W courses; after affirming a strong correlation between good assignments and good student writing (and vice versa), Freshman English developed an online archive of exemplary assignments and held teaching workshops on this topic; Nursing discovered that their students were not doing enough synthesis in literature reviews and made this a point of emphasis for future W courses; Electrical Engineering has similar plans focus more on interpretation of data in its lectures and through informal, write-to-learn in-class activities on data analysis; several departments that had no rubric for evaluating writing have since started using the rubric developed as part of this project; and some departments, such as Mechanical Engineering, realized that changing the writing process to have drafts due earlier, accompanied by student self-assessments, could improve the performance of all students, and especially those struggling with writing and revision.
- *These assessment projects contributed to the professional development of the faculty and graduate students involved.* Beyond supplying data to inform evidence-driven discussions of general education at the University of Connecticut, these studies served as modes of intensive faculty development for the participants. The 25 faculty and graduate assistants involved could focus, in a sustained way, on the writing of UConn students. To a person, participants affirmed that reading, scoring, and discussing the student writing would enrich their future teaching and professional growth.