Portfolio Assessment in Program Evaluation

Carol Siri Johnson, Norbert Elliot
New Jersey Institute of Technology

This paper describes the writing assessment program in the Department of Humanities at New Jersey Institute of Technology. Writing is an essential component of engineering education, and yet it is difficult to assess. At most universities, writing assessment centers on the evaluation of individual student ability for placement or gate keeping. At NJIT, we have elected to focus on using assessment to evaluate the impact of our humanities curriculum on our students, and on discovering ways to improve that curriculum. After providing an historical background, we will describe our program in its use of best paper and portfolio techniques. The conclusions that we will then draw reflect the iterative, information-based nature of the program—one that complies with national and professional accrediting agendas.

Background: A Culture of Assessment

How is writing viewed in the 21st century? As Roberta Camp of the Educational Testing Service has observed, writing is “a rich, multifaceted, meaning-making activity that occurs over time and in a social context, an activity that varies with purpose, situation, and audience and is improved by reflection on the written product and on the strategies used in creating it.” [1] In her longitudinal study of college writers, Marylin S. Sternglass observed that instruction in writing is critical to fostering an analytic stance in disciplinary courses, instruction that allows students to "understand the significance of ideas in their particular field to the level where they become able to question some of the assumptions of that field." [2] Within engineering education, Carolyn Miller has recently praised the communication-across-the-curriculum movement in her review of program outcomes established by the Engineering Accreditation Commission of the Accreditation Board for Engineering and Technology. That movement, she wrote, "aims to incorporate writing and speaking into every program and department across our campuses in order to demonstrate that these are not isolated subjects but rather arts that are relevant to every subject." [3] Hence, writing is a complex activity, itself a technology that enables in-depth learning in a wide variety of fields. Promoted across the curriculum, writing in the 21st century is far more than a skill: it is, rather, a vehicle for empowerment.

Beginning with our roles in the New Jersey Basic Skills Placement Test (NJBSPT)—a state-wide assessment system for placing new students in all state institutions from 1977 to 1994, employing nearly 43,000 yearly tests—NJIT faculty in the Department of Humanities have built a culture of assessment that is informed by such a construct of writing ability. Our construct—that writing is a cognitive activity, taught as a process, that increases learning across-the-curruculum—is applied to our university's General University Requirements in the humanities, a series of 18 credits taken by all undergraduate students from the freshman through the senior year.

After students are admitted to NJIT, our Department uses a combination of SAT Verbal scores, limited response sentence scores, reading comprehension scores (both derived from the NJBSPT instruments), and holistically-scored essays in order to place students in the freshman writing curriculum. That curriculum includes either a one-semester, three-credit course or a two-
semester, six-credit sequence, with parallel English as a Second Language courses. Once the freshman sequence is complete, students enter the following sequence: a two-semester, six-credit sequence of cultural history (required for all majors in the lower division); a one-semester, three-credit course in technical writing (required for some majors, elected by many others in the junior year); a one-semester, three-credit elective in literature, history, philosophy, or STS (Science, Technology, and Society); and a one semester, three-credit elective capstone in the humanities (required for all majors in the senior year). In that many of our students elect to take the junior-level elective in literature, history, philosophy, or STS at Rutgers-Newark, a campus adjacent to NJIT, we do not regularly assess those courses. On-going assessment is regularly undertaken in the freshman writing course (HUM 101, Writing, Speaking, Thinking), cultural history (HUM 211, The Pre-Modern World, and HUM 212, The World and the West), technical writing (Eng 352, Technical Writing), and the capstone seminars (a series of HUM 400-level courses).

This is the curriculum that we have elected to assess in order to help students become better writers. To aid us in the assessment, we have developed an outcomes assessment program that uses both best paper and portfolio reviews.

Program Assessment: The Use of Best Papers
Each semester, students submit best papers at the end of the courses in freshman writing, cultural history, technical writing, and the capstone seminars. Instructors inform their students that their papers will be evaluated by the stated criteria (see Figure 1 below) and, with their students, select the best papers for review. Instructors then meet at the end of each semester to review the papers. Historically, the papers have been read reliably by our experienced readers using a holistic assessment method; that is, the inter-reader reliability coefficients have always exceeded .6, the lowest minimum coefficient of correlation set by the Department to establish the inter-reader reliability needed to undertake further analysis.

Each semester, instructors—untenured and tenured, one-year appointments and lifers—gather together on Reading Day at the end of the semester to review papers on the six point scale. Readers use sample training papers to augment the scale and resolve discrepant readings (i.e., those that are not matching or adjacent) in order to award a score from 2 (the lowest) to 12 (the highest). At the conclusion of the reading, program coordinators create one-page summary sheets, such as the one shown in Figure 2, to use as a touchstone for discussion with instructors about the level of student achievement.

In the instance shown in Figure 2, instructors may be confident that they are evaluating as a community and scoring the best papers reliably (r = .79). In addition, it is clear that the scores are well distributed along the scoring matrix (from 2 to 12). And it is also clear that the scores across semesters are consistent, with no significant difference found across two administrations (t = .941, p = .1737). Thus, faculty and administrators may be confident that our assessment has verified that graduating seniors demonstrate proficiency in writing ability as measured by our assessment community.
SCORING LEVELS FOR WRITING IN THE HUMANITIES AND SOCIAL SCIENCES

Listed below are writing standards that are used to score analytic papers in the Department of Humanities and Social Sciences. This six-point scale may be used as established, objective criteria by which students and their instructors may come to consensus on the characteristics of effective written communication.

**Level 6**
- The entire paper works to develop a thesis.
- The thesis is supported throughout by specific arguments and evidence.
- Paragraphs are used effectively to develop each stage of the discussion.
- Transitional words, phrases and sentences are employed.
- Effective and appropriate word choice is made.
- There is evidence of extensive vocabulary.
- The paper is virtually free of grammatical or mechanical errors.
- Correct MLA or APA citation is used. Use of paraphrase and quotation is skillful.

**Level 5**
- With very few lapses, the paper develops a main idea throughout.
- The thesis is generally supported by specific arguments and evidence.
- Paragraphs are generally used effectively.
- Generally effective transition.
- Generally effective and appropriate word choice. Evidence of a good vocabulary.
- Very few grammatical or mechanical errors.
- Generally correct citation. Appropriate use of paraphrase and quotation.

**Level 4**
- The paper has a clearly recognizable thesis, but may occasionally stray from the topic.
- Specific arguments and evidence are marshaled in support of the thesis, though there may be significant lapses in this effort.
- Effective transitions are sometimes present.
- Word choice is functional, though sometimes reflects a limited vocabulary.
- Some grammatical or mechanical errors, though sentence boundaries and verb tenses are under control.
- An attempt to cite sources is made. An overall understanding of paraphrase and quotation is demonstrated.

**Level 3**
- There is some attempt to develop a thesis, but the paper is not generally effective in developing the main idea.
- There is some attempt to organize specific arguments and evidence, but this effort often breaks down.
- Effective transitions are seldom present.
- Vocabulary is limited.
- Grammatical and mechanical errors are fairly numerous, but there is no pattern of sentence or verb errors.
- Minimal grasp of citation style. Some confusion about the boundaries between the student’s own words and ideas and those of others. Confusion about the difference between paraphrase and quotation.

**Level 2**
- Only a minimal attempt at thesis development is made.
- Little success in marshaling specific arguments or evidence.
- No attempt to use transitional words, phrases or sentences.
- Very limited vocabulary.
- Numerous grammatical and mechanical errors, and a pattern of sentence and verb errors.
- Inability to cite correctly. Confusion about the boundaries between the student’s own words and ideas and those of others.

**Level 1**
- No recognizable thesis.
- No coherent use of specific arguments or evidence.
- No concept of transition.
- Extremely limited vocabulary.
- No sustained ability to control sentence boundaries or verb tenses. Numerous other grammatical and mechanical errors.
- Virtually no distinction between the student’s own words and ideas and those of others.
Program Assessment: The Use of Portfolios

While the holistic assessment of best papers supports the writing-as-process philosophy of the department—we assess only representative student work, thus decreasing the construct-irrelevance of timed tests of writing—our Department needed a way to evaluate analytically the writing of our students and to see more than a single sample of their writing.

In the fall of 2003, we therefore designed a sampling plan to give us a confidence interval of 90 to 95 %. In addition, we developed the criteria in Figure 3 to evaluate the portfolios, and we tailored each of the evaluative criteria to the courses we would assess. While the instructors were familiar with holistic scoring methods, the analytic method of scoring used in the portfolio evaluations was new, and thus the portfolios were scored by the program directors themselves.

[4]

Note: Mean in Spring 2003: 7.4137. There is no significant data in comparison across two semesters (t=.9431; p=.1737)
As Figure 3 shows, the Department was interested in assessing the following: the ability of students to think critically; the ability to students to demonstrate a global perspective in their analytic writing ability; a demonstration of drafting as part of our writing-as-process departmental pedagogy; the ability to cite texts as demonstration of their documentation skills and as evidence of their ability to use multiple texts to support their analytic writing; the ability of students to make oral presentations (evidence of which would be found in slides or handouts within the portfolios); the ability of students to work collaboratively (evidence of which would be found in drafts or in multiple-authored documents); and an overall portfolio score, a holistically-derived evaluation. In relational terms, we believed that the ability to demonstrate critical thinking, to exhibit global perspective, to use writing-as-process, to cite sources, to deliver oral presentations, and to work collaboratively were associated with the ability to write in
an academic setting. The lower the levels identified, the more work would be needed to strengthen the curriculum; the higher the levels identified, the more the Department could be confident that the curriculum was meeting the curricular goals embedded in the portfolio assessment instrument. Figure 4 presents a summary of our analytic reading, again achieving inter-reader reliability ratings of .6 or above on each of the traits measured.

Figure 4. Summary of Portfolio Reading, Fall 2003 (Freshman Composition N=58; Cultural History N=19; Technical Writing (N=48); Capstone Seminars N=46)

As Figure 4 indicates, critical thinking (ranging from 8.28 in cultural history to 9.22 in the capstone seminars) and overall score (ranging from 8.55 in cultural history to 8.93 in the capstone seminars) received the highest ratings. Indeed, these high scores reflect the high scores regularly recorded in the best paper readings. The analytic scores from the portfolio reading, however, reveal room for curricular improvement. For example, in our cultural history course sequence, students did not provide evidence of multiple drafts of papers in their portfolios, a problem captured in the mean score of 5.17. Comparatively lower-level oral presentation scores were recorded both for freshman writing (6.6) and the capstone seminars (6.39). Collaborative work also received comparatively low scores in freshman composition (6.3) and cultural history (6.1). Our assessment cycle for the year 2003-2004 was concluded, but our curricular work had just begun.
Conclusions: An Adherence of to Standards, a Culture of Humility, and a Process of Iteration

The Middle States Commission on Higher Education is explicit about outcomes assessment in the 2003 accreditation guidelines. Not only must outcomes of learning be established and assessed, but the assessment process must demonstrate that the results have been used to improve teaching and learning. In identical fashion, the 2004-2005 criteria established by the Accreditation Board for Engineering and Technology require assessment processes that show that results are used to further program improvement. In short, goal setting and outcomes assessment without follow-up are insufficient in higher education for the 21st century. National standards tell us that evidence must be given that the assessment results are being wisely used to improve student learning. The age of accountability, begun in the 1970s, had taken a sharp turn in 2004.

In such an environment, we believe that it is best to adopt a framework of humility in designing, implementing, and using information gained in programs of outcomes assessment. In the Department of Humanities at NJIT, we realize that a single sample of writing ability taken on a single day yields limited information. We seek to extend our evaluation program beyond cut scores and notions of passing and failing and also acknowledge the human side of assessment: students may not have given their best work in a single paper or included all of their drafts in a portfolio; instructors may not have conveyed well the assessment criteria; and analysts may favor one interpretation over another. Key in such an environment, then, is to acknowledge that it is best to look for trend information, for patterns in student learning that will inform formative curricular decisions. Absolute proof may be said to exist as a result of research methods in the natural sciences, but there is no indication that those methods are possible in educational research. As Martin Hollis has written, "we have found no single and commanding analysis of causal explanation in the philosophy of the natural sciences which social scientists are bound to accept." [5]

An example of the kind of humility we are advocating may be found, again, in our analysis of Figure 4. As the chart indicates, critical thinking and overall score received the highest portfolio ratings (e.g., scores above 7), ratings that we regularly record as a result of the best paper readings (e.g. those shown in Figure 2). But with years of outcomes evaluation experience behind us, we knew that this similarity would exist. Best papers evaluated holistically implicitly acknowledge a student's ability to think critically, and so we expected the portfolio evaluation of critical thinking and overall score to reflect similar achievement levels. Candidly, it would have been a simple matter to stop there, report the results, and—with full hubris—await administrative acknowledgement of work well done. Yet we wanted more for our students. We suspected, for example, that writing-as-process was working well in freshman composition and in technical writing, but we knew little about the adoption of that pedagogy in cultural history and in the capstone seminars. While we were pleased to see high scores in the capstone seminar (8.8), the low scores in cultural history (5.17) gave us pause. Similarly, we knew that our techniques were limited in evaluating evidence of oral presentations by looking for slides or notes, mere written products of complex performances. Similar, product-oriented flaws were also clear to us as we sought to evaluate collaborative work.

With these results in hand, we can respond to the need for further program development. As a result of our assessment, our cultural history coordinator has begun to institute a writing-as-process pedagogy into the courses in which a persuasive course paper is proposed early in the
semester, the proposal is approved, and the final paper is then drafted into a final copy with the instructor. We also realized that the academically-oriented criteria used to evaluate documented essays in freshman composition, cultural history, and the capstone seminars were not completely valid in evaluating the non-academic discourse of technical writing. At present, new criteria are being designed for the technical writing course to include attention to visual clarity, presentational logic, audience-awareness, appropriate use of voice, and editing competency. (A spirited debate, we note, surrounded the issue of documentation of sources—a criterion which we will now include in technical writing because a consistent attitude toward academic integrity is critical to student development.)

In sum, our assessment program in the Department of Humanities provides a basic structure for sustainable curriculum improvement. By assessing programs (not the students—that is the job of the instructors) our community can initiate changes that will be used to continually improve the program. At NJIT, we plan to pursue a number of strategies for refining our assessment metrics: our goal is to discuss, debate, and present assumption of core competencies that drive assessment, and (in turn) to revisit, refine, and clarify the construct validity of the existing predictor variables. Our strategies will include engaging in face-to-face and electronic discussions among faculty about the design of programmatic standards and the methods used to ensure that they are being met. And, as always, we will communicate our curricular goals to those most impacted by our decisions: the students. When will the process be complete? Never.

NJIT is devoted to the education of engineering students and we are one of the most diverse universities in the nation—nearly 14% of our students tell us that English is not their first language. Capturing this diverse student population—in 2003 an entering class of 44% White Non-Hispanic students, 23% Asian/Pacific Islander students, 11% Hispanic students, and 8% Black Non-Hispanic students—requires careful sampling plan design, a process that unifies statistical and humanistic impulses. If we are to design programs that incorporate a continued commitment to access for diverse student populations, we need to attend to the ways that our programs foster success for all students. At NJIT, the Department of Humanities is committed to ensuring the success of our diverse student population in an ever-changing multicultural and multinational world. Assessment, often considered a dreaded educational act, can become a vehicle for ensuring ongoing democratic realization within the academy.

References


[4] For a description of the analytic scoring methods we followed—as opposed to holistic or primary trait methods—see Alan C. Purves, Thomas P. Gorman, and Sauli Takala, "The Development of the Scoring Scheme and


**Biographical Information**

CAROL SIRI JOHNSON is the director of the undergraduate technical writing service course and the assistant director of the Master of Professional and Technical Communication program at New Jersey Institute of Technology. She has a B.A. in Studio Art from Mount Holyoke College and a Ph.D. in English from The Graduate Center of the City University of New York. She worked as a technical writer in the computer industry for six years before coming to NJIT. Her areas of research are assessment and the history of technical communication.

NORBERT ELLIOT is professor of English at NJIT. His most recent book, *On A Scale: A Social History of Writing Assessment,* is due in 2005 from Peter Lang.

**Acknowledgements**

This paper is part of an on-going research effort by the program directors of the Department of Humanities at NJIT: John Coakley (Freshman Composition), Burt Kimmelman (Cultural History), and Nancy Steffen-Fluhr (Capstone Seminars). Nancy Coppola directs the MS in Professional and Technical Communication and, with Carol Johnson, is responsible for the development of technical writing. Robert Lynch, the Chair of the Department, is the longest-serving and most experienced member of our research team. Vladimir Briller, Director of Outcomes Assessment in the Office of Institutional Research, is our expert in sampling plan design and data analysis.