

Writing Assessment: An Evaluation Paradigm

Harvey S. Wiener

The pressure for accountability in postsecondary education continues to build. Recent reports from the National Institute of Education and the Association of American Colleges indict the academy for inattention to assessment, which both documents insist is a key indicator of an institution's health. It is hard to fault the conclusions of these reports. Although I do not wish to minimize some of the strides we are making in large-scale assessment efforts—the innovative programs at Alverno College, for example, the Exxon funded consultation and evaluation project of the Council of Writing Program Administrators, the Student Outcomes Project of the University of Tennessee, to name just a few—at the classroom level, where teachers must judge what students have learned or need to learn, we have few paragons. For humanistic studies, accountability is particularly complex and hard to manage. However, it is not likely that we can shrug our shoulders in puzzlement much longer as the public continues to demand systems for measuring achievement in all quarters. Either we face the challenge of developing our own tests, or we submit to inferior schemes by testing specialists or their surrogates on the fringes of the academic discipline.

Over the last decades teachers of English composition have wrestled with the issue of testing student writers and have devised an assessment system whose history is a notable model for the profession at large. What I wish to suggest here is that the development of current tests for writers is a paradigm for how other liberal studies faculties can shape valid assessment programs that reflect the important tenets of their disciplines.

In early efforts to assess writers, the multiple choice exam was the most ubiquitous testing format. Having thought little about the relation between evaluation and instructional goals, we listened to outside testing experts, who insisted that with a series of questions on usage, grammar, and syntax, the darkened boxes on answer sheets would provide "predictive validity," a crystal ball for students' writing skills. We could determine entry-level proficiency, for example, without asking students to write a word. But it wasn't long before many teachers of writing rejected assessment measures that did not demand the construction of sentences. Thus, the "writing sample" test grew in prominence.

(In a 1983 survey of member institutions of the National Testing Network in Writing, 69% of the respondents reported that they assessed writing with a writing sample. Many of these are locally developed and scored tests.) The new model departed from accepted testing principles: Writing samples took too much time, cost too much money, and taxed teaching faculties quite severely. Yet the move from a short answer test to a sustained written response was an important positive step. The act of writing calls for language and syntax specific to individual thinkers. Testing students on weaknesses in someone else's prose had little utility, even though information to be gathered from the tests was inexpensive, could be easily quantified, and might imply students' later successes. With a writing sample, English teachers drew upon their knowledge of current composition theory, sound teaching practices, and writing assessment models developed by national testing agencies in order to devise an acceptable evaluation format. To judge writing, you looked at writing, and that was that.

Despite the advance to a new type of test, many writing teachers clung nonetheless to the values of the older model, which viewed writing as a technology. Thus, we simply counted errors—seven or eight mistakes, and the paper failed. Error counting paid no attention to the nature of thinking in writing and offered not much improvement over the machine-scored test. Examining the surface features of written language is certainly the easiest way to attend to a piece of prose, but it is also the least important in telling what students need to know and learn.

In the next phase of the paradigm I am outlining, teachers of writing moved to a much more complex and demanding view of assessment, one based upon thought *and* expression in the writer's prose. To look at the contents of a piece of writing produced under restrictions of time and format and demanding rapid evaluation for placing students into appropriate courses or programs, evaluators had to develop a scoring system that would lead to sound judgments. With research, hard work, and much experimentation several such systems have emerged over the years. We now have various methods of looking at an essay besides counting its mistakes; these methods all insist that what the writer says and how the writer says it are the true, integrated territory for appropriate evaluation.

What I am suggesting is that in a humanistic discipline like writing, the concerted efforts of faculty can lead to evaluation instruments and designs that are both manageable and responsive to the complexity of the task awaiting judgment. I admit that I have oversimplified the development of the paradigm. It is not linear, and it does not always run on a smooth track. Several institutions still assess writing without asking

students to write. Some English teachers have closed their eyes to the need for valid writing assessment models and the recent progress made in devising them. Others still argue for machine-scored "objective" tests. And certainly supporters and detractors of the popular scoring systems are waging their wars to this very moment. (Witness the December 1984 issue of *College Composition and Communication*, a battleground given over exclusively to testing concerns for the first time.) Still, the important consideration here is that by 1985 many institutions have moved from writing assessment rooted in quantification to writing assessment rooted in informed and carefully monitored personal responses. For any of the methods of evaluation that enlightened writing faculties now used in grading a student's prose, teachers collaborate on goals and standards, rate papers in groups and under supervision, and adapt their individual perceptions of quality writing to the perceptions of their colleagues. These collaborative efforts at assessment have developed through continuous discussion and analysis at individual campuses, at local forums, and through national publications and conferences.

Ongoing conversation with colleagues has helped writing faculty move away from quantification and idiosyncratic personal judgments of a writer's prose to personal judgments based on collaboration and shared visions. By creating conditions for making defensible judgments, we have established a method for assessing the kinds of complex performances so typical of humanistic study. This is the model to lay alongside the data-driven, number-haunted statistics that dominate educational evaluation. Teachers of writing have nurtured a paradigmatic alternative to what is frequently mindless quantification. The current state of writing assessment asserts that a concerned, interactive professoriate can establish commonalities for judgment and can anchor them in rational thought and liberal learning. About this kind of evaluation many will sneer "soft data" and "subjective." But soft data in this case is human data, the record of minds thinking, judging, and evaluating writing, an act that is itself quintessentially an effort of thought, judgment, and evaluation. Only human minds can evaluate another mind's thoughts.

And if this is subjectivity, writing teachers have made the most of it. The NIE report complains that with new technologies we risk the loss of "human and subjective elements from both assessment and learning" and that "We do no great service to students by pretending that subjectivity does not exist in judgments of performance in the world. . ." Undeniably, objective quantification is useful; when teachers of writing collate the numerical representations of their judgments of student writing and relate them to each other and to the public, they are

in debt to numbers and statistics, charts and graphs. But the collaborative procedures writing teachers have evolved over the years demonstrate sensitive evaluation efforts based on the most important commodity in the humanistic enterprise, the mind in intellectual negotiation with peers.

Local assessment models initiated and developed by faculty are a relatively new and welcome addition to the educational enterprise. Unschooled in statistics and psychometrics, many humanists believe that they cannot influence an already entrenched testing establishment. They cringe at the apparent *reductio* of numbers, validations, statistical significances, *n*'s, and cohorts. They have looked the other way as professional testing agencies, and not classroom teachers, have developed entrance, proficiency, and licensing exams for undergraduates and graduates. They have permitted legislatures to promulgate competencies and have allowed state governments to mandate and oversee the testing of students' skills. The Florida State Department of Education, for example, has created the College Level Academic Skills Test for all students in the state, and has prescribed the number of pages to be written each week in writing classes. Similar efforts await action in other states. Certainly these are faculty, not legislative prerogatives. Yet when faculty do not act, they abdicate, and in the end, must be prepared to serve what may be an uninformed, unsympathetic, and unwelcome ruler.

About assessment, evaluation, and accountability, it seems to me that the AAC report is right: "In a society where survival and growth are often the only tests of virtue, colleges and universities have paid too little attention to an assessment of their performance." This is a gauntlet that educators must seize quickly. The evolution of writing assessment as a force in educational testing is a promising model for those who seek to define evaluation procedures that grow from sound academic principles and not simply from the expediencies of budget offices, campus testing services, or legislative mandates. Only from the imperatives of subject matter and the teachers committed to it grow the designs for accurate and valid assessment.

