16

MORE WORK FOR TEACHER?

Possible Futures of Teaching Writing in the Age of Computerized Assessment

Bob Broad

[Household] labor-saving devices were invented and diffused throughout the country during those hundred years that witnessed the first stages of industrialization, but they reorganized the work processes of housework in ways that did not save the labor of the average housewife.

—Cowan

In her book *More Work for Mother: The Ironies of Household Technology from the Open Hearth to the Microwave* (1983), Ruth Schwartz Cowan presents a feminist history of modern household technology. As the title of her book emphasizes, her argument is that the hundreds of "gadgets" invented with the purpose of easing the labor of "housewives" achieved the net result of dramatically increasing the quantity and range of tasks for which women were responsible in the American home. For example, when the wood-burning stove replaced the open hearth as the home's source of heat and the cooking apparatus, men and children (the family's collectors of wood fuel) had much less work to do because the stove consumed far less wood than did the open fireplace. However, the stove made it possible, and shortly thereafter obligatory as a sign of her family's increasing social status, for a woman to cook a much wider range of foods, often all at the same time, and most often with no help from anyone else in her household.

Likewise with the vacuum cleaner (which changed cleaning carpets from a semiannual family task to a weekly solo task) and the washing machine (which allowed soap producers to convince women that stained clothing was intolerable). In every case, the new technology briefly fulfilled its labor-saving promise before the social system—within which women's work was understood, negotiated, and shaped—quickly and substantially increased its expectations for women's household labor. The dramatic result of this dynamic was that numerous social observers during the nineteenth and twentieth centuries commented

on how American women, surrounded by helpful technology, invariably appeared pale, exhausted, harried, and sick. One lesson Cowan wants us to draw from her analysis is that technological advances must be analyzed and acted upon with careful attention to the social, cultural, and political *systems* within which they will play out in people's lives. Often, technology will deliver very different results as it plays out *in the social and political system* from what its designers intend or predict.

As an admirer of Cowan's history of technology, I immediately thought of *More Work for Mother* when in fall 2003 I drew from my mailbox a postcard from the Educational Testing Service (ETS). The postcard's rhetorical purpose was to persuade me (presumably along with everyone else on the National Council of Teachers of English mailing list or those interested in writing assessment) to visit the Educational Testing Service booth at the council's annual conference with the specific purpose of learning more about—and subsequently buying—an ETS product called Criterion.

As readers of the current book are probably already aware, ETS Technologies, Inc. is "a for-profit, wholly-owned subsidiary of ETS" (Burstein 2003, 119), and Criterion Online Writing Evaluation is a computer program that claims to evaluate students' writing. Two distinct appeals on the postcard from ETS attempt to persuade teachers of writing to include Criterion in their teaching practices. The most direct appeal shrewdly targets the topic on which writing teachers are most sensitive and vulnerable: time starvation. The postcard generously offers that "Criterion™ gives teachers what they need most . . . time to teach." Thus we classroom teachers of writing are encouraged to outsource to Criterion our evaluations of students' writing, the part of our job that, following the logic of the postcard, takes valuable time away from the *teaching* of composition. In a moment, I will examine the implicit belief that writing assessment takes us away from teaching writing. First, however, let me offer some personal context for my analysis of the second, more diffuse, appeal made by ETS on behalf of Criterion.

In the late 1980s, I was teaching high school English in Washington, D.C. and resenting the influence of testing—primarily ETS testing in the form of Advanced Placement and SAT exams—on the learning and teaching in my classrooms, especially the learning and teaching of writing. One evening I stopped into a bookstore near Dupont Circle and discovered a volume that permanently changed my professional life. The book was *None of the Above: Behind the Myth of Scholastic Aptitude* by David Owen (1985). I recommend this book to everyone concerned

with the effects of testing on education. Owen started out researching ETS in a quiet, journalistic fashion. Along the way, he was so horrified and outraged by the secretiveness, deceptions, and arrogant disregard for students and teachers demonstrated by those he dealt with at ETS that the book ended up as a blistering critique of the educational, political, and economic functions of ETS.

This is the right time to say that, as I've matured, my view of ETS has moderated to an extent. It has become clear to me that people at ETS are smart and dedicated, and furthermore that many of them do care about education. I have even met one ETS researcher in the past few years who shows a spirit of genuine intellectual inquiry (as opposed to the typical relentless ETS sales pitch) in his presentations at the Conference on College Composition and Communication. So my view of ETS is no longer as narrow or as fiery as Owen's. Nevertheless, it is clear to me that the *educational effects* of ETS products and services create serious educational difficulties and obstacles for students and teachers, difficulties I hope are unintended and unforeseen by the people of ETS but that teachers of writing must nevertheless vigorously oppose.

Now back to the postcard. The second appeal on the card has to do with the general character of the professional relationship between ETS and classroom teachers. Above, I have sketched my deeply skeptical, often angry, analysis of that relationship based on my reading of Owen and my twenty-five years teaching literacy to students from prekindergarten to doctoral studies. The postcard attempts to project exactly the opposite picture of that relationship under the headline "Working Together to Advance Learning" and "Listening. Learning. Leading." "At ETS we are committed to understanding the demands of the classroom by forming partnerships with educators. Your requirements drive us to develop products and services that help advance learning . . . . How long does it take you to evaluate an essay? Instantly . . . using Criterion™ Online Writing Evaluation."

In pitching Criterion as a solution to the brutal demands on writing teachers' time, ETS shows there is some truth to its claim to "understand the demands of the classroom." But in order to legitimately claim that they "listen," "learn," and "form partnerships with educators," ETS would have to do more than understand (and exploit) those classroom demands. They would also have to demonstrate understanding of the educational goals and values that shape classroom activities, and they would have to respond in some substantive way to teachers' concerns, expressed forcefully now for decades, about the detrimental effects on

U.S. education of ETS products and services. We are still waiting for such a listening and learning response from ETS, and we are still waiting for such a partnership.

The history of technology in general and of ETS's impact on teaching writing in particular make obvious that the time-saving claims ETS makes on behalf of Criterion are disingenuous at best. Cowan would warn that it is highly unlikely that outsourcing evaluation to ETS will result in teachers having more time to teach writing. To the contrary, writing teachers in institutions that purchase Criterion would almost surely be assigned more students or more classes, or both. Even more sinister than the implications for teacher workload and time is the quality or character of the educational impact programs like Criterion would have on the teaching of writing. Better than anyone in the world, the good people of ETS know that assessment drives instruction. A more candid motto for Criterion would be: "Teach your students to write like machines for a reader who is a machine."

PRESERVING THE PLACE OF ASSESSMENT WITHIN THE TEACHING OF WRITING

Let us now return to ETS's implicit claim that time writing instructors spend on assessment is time taken away from teaching. It is not difficult to see why a corporation that makes its money from the outsourcing of assessment would promote this view. The crucial question is whether students and teachers of writing, and the general public, ought to accept such a view and endorse it by purchasing products that help to separate teaching from assessment.

Evaluation of writing holds an undeniably murky and ambiguous place in the hearts of most writing teachers. Many of those teachers openly dread evaluating their students' writing, chiefly because it requires a tremendous investment of time and effort, yielding often dubious pedagogical benefits (see Belanoff 1991; Haswell, chapter 4 in this volume). However, this bleak scenario is not the only one possible for teachers of writing. In fact, many of us do some of our highest-quality teaching when responding to and evaluating our students' writing. And the core argument of one of our profession's most important recent books is that teachers of writing need to reclaim assessment as a crucial, powerful, and rewarding part of the process of teaching and learning writing.

Brian Huot's (*Re)Articulating Writing Assessment* (2002) urges teachers of writing to lay claim to evaluation and use its power to drive the best possible teaching and learning of composition. "Assessment can and

should be . . . an important and vital part of the effective teaching of writing. One of the main goals of this book is to establish the importance of assessment to the teaching of writing and to connect the teaching of writing to what we now call writing assessment" (11).

Huot's insistence that composition instructors embrace and control the design and implementation of assessment as an integral part of their teaching practice obviously and directly contradicts the enthusiasm for outsourcing assessment evident in the postcard from ETS. But Huot goes even further than making claims for the pedagogical importance of understanding evaluation as part of teaching. Citing Beason, Huot also points openly to the power dynamics and ethics of the relationship between teaching and assessment. "Our profession's abandonment of assessment as a positive practice and its adoption of negative conceptions of assessment as punitive and counterproductive to fostering literate behavior in our students cannot but continue to put us in a position of powerlessness, while at the same time putting our students and programs in peril. To come to a new understanding of assessment is to not only become conscious of its importance, power, and necessity for literacy and its teaching, but also to understand assessment as one of our ethical and professional responsibilities (Beason 2000)" (13). Taking up the power inherent in evaluation is not only our pedagogical responsibility but also our political responsibility.

To cap off his proposed transformation of the relationship between evaluation and teaching writing, Huot even argues that teachers need to teach their students how to assess their own and each others' writing as part of the crucial set of rhetorical skills students need to be successful writers. "Being able to assess writing quality and to know what works in a particular rhetorical situation are important tools for all writers" (2002, 70).

As we contemplate the possible futures of teaching writing in the age of computer-assisted writing assessment, we—teachers, students, administrators, and the public that funds and benefits from our work—need to choose between a vision of literacy learning like Huot's that includes, embraces, and enhances the educational power of assessment or the ETS sales pitch suggesting that teachers and students will be better off when students' writing is evaluated by a computer instead of by themselves, peers, or teachers.

TEACHING WRITING IN THE AGE OF COMPUTER-ASSISTED WRITING ASSESSMENT

Current thinking in our field holds that outsourcing assessment comprises a seriously damaging loss to the teaching and learning of writing.

Chapters in the present volume (including this chapter) also view the use of computerized writing assessment as potentially damaging to students' rhetorical development. Here, I aim to develop further the argument against using computerized evaluation in teaching writing and to discuss what the history of technology and democracy have taught us about how to win that argument. I will also speculate methodically about what our future(s) might look like if we lose the argument and university and college administrators outsource writing assessment to dealers such as ETS, Vantage Learning, and Knowledge Analysis Technologies.

Though I have what I consider a satisfactory knowledge of the technical workings of products such as Intelligent Essay Assessor, Criterion and IntelliMetric my focus in this chapter is philosophical and strategic rather than technical. To help teachers of writing protect the integrity of their profession, I will apply lessons learned from the history and philosophy of technology to the emergence of new-generation computerized assessment, and I will apply lessons learned from past struggles between technocrats and the wider public for control over how technology is distributed and used in society. For support in these efforts, I will begin and end by looking to the work of Andrew Feenberg.

Feenberg's 1991 *Critical Theory of Technology* initiated his argument (carried forward in his subsequent work, to which I will turn later) regarding how societies might shape uses of technology for the common good as discerned through democratic processes. Feenberg insists that the typical dichotomy between technophobic and technocratic viewpoints will fail to serve this project. Instead, he suggests that we stay alert to technological developments and make, as a democratic society, well-reasoned decisions regarding how to handle those developments.

For the specific purposes of considering how teachers of writing might best respond to the emergent assessment technologies, I found Feenberg's thoughts on the interplay between technology and understandings of human capabilities especially helpful. "Roughly formulated, the problem concerns the similarities and differences between human thought and information processing. To the extent that similarities can be found, computerized automata can replace people for many sophisticated purposes. To the extent that differences are found, greater philosophical precision is introduced into the notion of human thinking, clearly distinguished from manmade simulacra" (96–97).

What I, as a teacher-scholar of composition, take from Feenberg's analysis is that insofar as my assessment processes match what a

computer can do, it may be appropriate and helpful to use the computer in evaluating students' writing or shift my teaching to address other needs. And where my teaching practices offer students value that computers cannot reproduce, I should not only productively focus my professional energies for the sake of my students, but also strive to understand myself and my profession with greater insight and "precision."

One thing that makes the history and politics of technology exciting is that no one really knows where technological developments may take us. So while we might take heart from the humanistic undertones of Feenberg's approach to technology, not everyone is so kindly disposed toward the human side of the equation. Take, for example, Ray Kurzweil. In *The Age of Spiritual Machines: When Computers Exceed Human Intelligence* (1999), Kurzweil frankly and with tremendous enthusiasm predicts that humans will soon be reduced to the status of technology's mainly extraneous caretaker. "In the second decade of the [twenty-first] century, it will become increasingly difficult to draw any clear distinction between the capabilities of human and machine intelligence. The advantages of computer intelligence in terms of speed, accuracy, and capacity will be clear. The advantages of human intelligence, on the other hand, will become increasingly difficult to distinguish" (4). Kurzweil's extended analysis makes clear that, given sufficient enthusiasm for and faith in the rapid development of artificial intelligence, Feenberg's formulation could leave humans with nothing to offer, and nothing to do, that computers can't do better, quicker, and cheaper.

Luckily, the new generation of computerized assessment technologies, while undeniably impressive from the standpoint of artificial intelligence and language processing, leave human teachers of writing with plenty to do. More important, those technologies can help us better identify and understand what we human teachers of writing do best, especially writing assessment. For now, Feenberg's analysis still holds promise. But the future of our profession depends on how we understand and represent to ourselves and the general public what rhetoric and rhetorical instruction are.

PREDICTIONS AND QUESTIONS FOR THE FUTURE OF RHETORICAL ASSESSMENT IN "THE AGE OF SPIRITUAL MACHINES"

Fortunately, our most robust definitions of rhetoric promise to hold computerized evaluations at bay for some time. Consider James Berlin's 1996 description of rhetoric's dynamic and multidimensional processes.

"Thus, in composing or in interpreting a text, a person engages in an analysis of the cultural codes operating in defining his or her subject position, the positions of the audience, and the constructions of the matter to be considered. . . . The reader must also engage in this dialectical process, involving coded conceptions of the writer, the matter under consideration, and the role of the receiver in arriving at an interpretation of the message" (84). In Berlin's description, reading and writing involve complex interpretations of cultural codes. Berlin drills deep into rhetoric to find the most sophisticated and nuanced elements of the process.

In the world of computerized writing assessment, the usefulness of a definition of rhetoric like Berlin's is that evaluation software doesn't even begin or claim to assess these cultural and intellectual capabilities (see also Ericsson, chapter 2 in this volume). The theory of rhetoric underlying computerized evaluation is relatively rudimentary and reductive. Designers of such products as Intelligent Essay Assessor forthrightly admit that they are simply incapable of (and uninterested in) assessing rhetorical abilities. "Mr. Landauer says it [Intelligent Essay Assessor] is not intended to be used for English-composition or creative-writing assignments, in which a student is being graded more on writing skill than on knowledge of a subject. The essay assessor works best on essays assigned to check students' factual knowledge in such subjects as history, political science, economics, and the sciences" (McCollum 1998, A38; see also Landauer, Laham, and Foltz 2003).

Even the spokespeople for ETS's e-rater (the "scoring engine" for Criterion; see Burstein 2003, 119), which is designed and marketed (as we have seen) specifically for the assessment of rhetorical abilities, admit that their product cannot assess the stylistic and intellectual merits of texts. "[Richard] Swartz [of ETS] emphasized the modest goal of computerized scoring: to judge the structure and coherence of the writing, rather than the quality of the thoughts and originality of the prose. In college, he said, professors grade the development of ideas, while essay-rating computers 'are better suited to judgment about more basic-level writing'" (Matthews 2004).

The point here is not the limitations of a particular computerized evaluation system, nor even whether we agree with Berlin's definition of rhetoric. The point, following Feenberg, is that understanding what artificial intelligence can do should and will shape our conception of what human intelligence can do. In this way, mechanical assessment promises to help us clarify and refresh our understanding of what we do

when we teach and assess writing by distinguishing what humans can do from what computers can do.

PREDICTIONS

The history of technology suggests that the continued growth and evolution of artificial intelligence will privilege what humans alone can do and commodify (and thereby devalue) what computers can do. Based on the capabilities and limits of artificial intelligence in relation to the practice of teaching and assessing writing, we should expect that computerized assessment will lead us to privilege several specific features of writing instruction.

- *Rhetoric* (see Berlin 1996) as a process so complex and multiply context-dependent that only human beings can successfully perform and analyze it
- *Feeling* (curiosity, humor, irony, pleasure, desire) in evaluating writing
- Human *relationships* in the learning and teaching of writing: teachers and students working, negotiating, and creating knowledge collaboratively
- *Diverse kinds of readings*: poetic, perfunctory, generous, mean-spirited, imaginative, critical
- *Validity* and *educativeness* (see Wiggins 1998) of evaluations

Meanwhile, by handling the following aspects of writing instruction competently, computerized assessment will commodify them, and so lead us to devalue them as processes a mere machine can perform.

- The composition and evaluation of *standardized timed impromptu essays* and essays written chiefly to show *content knowledge*
- Quick, cheap, quantitative *grading* or *scoring*
- Numerical *agreement* ("reliability") as a feature of multiple evaluations

PROMISING AVENUES OF INQUIRY

In addition to shifting how we value different elements of writing assessment, emerging evaluative technologies also raise interesting new questions in the field of teaching and assessing writing.

- Given a choice between human and computer evaluation of their writing, which will students choose, and why? (In their preliminary inquiries, Baron 1998 and Foltz 1998 came up with conflicting answers to both the "which" and the "why" questions.)
- What do students learn about writing when their performances are evaluated by a computer? How does the expectation of computerized grading shape students' writing processes and products?
- As more computerized assessment programs enter the marketplace, how will they compete against one another? What features will distinguish one computerized evaluator from another? Will the effort to compete through emphasizing such differences undercut mechanized assessment's claims to objectivity and neutrality?

FIGHTING TO PRESERVE HUMAN WRITING ASSESSMENT

In spring 2004, the Conference on College Composition and Communication issued its "Position Statement on Teaching, Learning, and Assessing Writing in Digital Environments" (2005). The statement presents a number of thoughtful observations and guidelines for teachers and administrators of writing programs using digital technology. Near the end of the document, the position statement addresses computerized writing assessment. Under the heading "A Current Challenge: Electronic Rating," the statement makes this unambiguous assertion: "Because all writing is social, all writing should have human readers, regardless of the purpose of the writing" (789).

For those of us committed to rhetorical education, this bold, clear statement from the conference in support of human readers is very welcome. To prove effective in protecting writing classrooms from efforts to outsource assessment, however, the statement will need further support and development. For starters, we will need to follow the advice of the conference statement itself: it states that decisions about teaching and assessment practices must be justified with direct reference to learning goals or outcomes.

> As with all teaching and learning, the foundation for teaching writing digitally must be university, college, department, program, and course learning goals or outcomes. These outcomes should reflect current knowledge in the field (such as those articulated in the WPA [Writing Program Administrators] Outcomes Statement), as well as the needs of students,

> who will be expected to write for a variety of purposes in the academic, professional, civic, and personal arenas of life. Once programs and faculty have established learning outcomes, they then can make thoughtful decisions about curriculum, pedagogy, and assessment. (786)

The conference position statement aptly suggests the WPA Outcomes Statement (Harrington et al. 2001) as just the sort of outcome statement that can best guide our assessment decisions. Yet in rejecting computerized evaluation in favor of human evaluation, the conference statement does not support or justify this particular position with reference to the WPA statement or other specific learning outcomes.

The simultaneous emergence of commercial computerized assessment and the bold but as yet unsupported stance of the Conference on College Composition and Communication statement in favor of human assessment creates the perfect opportunity for those in our profession to articulate how human writing assessment better supports our desired outcomes for rhetorical education than does mechanical evaluation. Therefore the area of future struggle will be over which outcomes (that is, which kinds of learning, skill, and knowledge) are valued most highly. If we sincerely believe it—and I hope we do—we need to emphatically argue exactly how human instructor-evaluators provide superior educational experiences over the unarguably cheaper, faster computerized evaluation.

Fortunately, the WPA Outcomes Statement provides a vision of and mission for rhetorical learning, teaching, and assessment that strongly supports human evaluation. Perhaps the greatest threat to the century-old project of standardized writing assessment is the fact that rhetorical processes are highly varied and context-sensitive. Because standardized writing assessment has always relied on a theory of classical psychometrics (now outdated in the field of psychometrics), it has always emphasized standardization and consistency over the variation and difference that are the marks of rhetorical exchange. The Outcomes Statement boldly foregrounds the need for our students to learn to respond to "the needs of different audiences" and "different kinds of rhetorical situations." Note that computerized assessment's distinguishing feature is decontextualized and generic (i.e., standardized) rhetorical tasks, and its main point of pride is the sameness and consistency of the scores it awards to student texts. The Outcomes Statement helps clarify that the overwhelming uniformity inherent in mechanical assessment undermines our efforts to prepare students to compose, assess, and succeed in complex and varied rhetorical scenarios.

The Outcomes Statement also highlights such complex skills as selecting, evaluating, and using sources; thinking critically and creatively; and mastering various modes, phases, and strategies of the composing process. These are yet more areas in which artificial intelligence has not even claimed the right to encroach on human teaching and assessment, much less proven itself worthy to do so. So this is the terrain on which we will struggle to preserve and privilege human judgment in teaching writing.

In his 1999 *Questioning Technology*, Feenberg points to historical examples of people rejecting technocratic control of technology in favor of democratic control of technology. In the cases of AIDS drug treatments, computer networks, and various environmental threats, people saw technology playing out in society in ways they determined harmed their values and their goals, and they organized politically to change their course.

Luckily, those of us concerned with teaching and assessing writing do not have to look far for examples of how we have successfully supplanted destructive assessment practices with constructive ones. Both in 1943 and in 2004, on the eve of the unveiling of SAT II, the Educational Testing Service bowed to pressure from writing teachers and, contrary to its best technocratic and classical psychometric judgment, included an actual writing sample in its assessments. Commentators like ETS's Hunter Breland (1996) have fumed over the ignorance and stubbornness of writing teachers in insisting on making people write when assessing writing ability, but writing teachers have nevertheless (so far) carried the day.

In our current efforts to understand computerized assessment and determine its appropriate place in the realm of rhetorical learning, we can follow these historical examples from within and outside of rhetoric and composition. First, as McAllister and White argue in chapter 1 of this volume, we have the responsibility to educate ourselves about the features and implications of various mechanical-assessment applications. Next, we have the solemn responsibility to study and predict the impact on rhetorical learning of these various applications. If we, as professional educators, determine that a particular use of artificial intelligence helps students and teachers meet established learning goals, then we should support and invite that use of technology. Where we determine that use of computerized evaluation would trivialize and denude rhetorical instruction and experience, we must fight it and prevent it from being used.

Victory in this struggle will depend on our ability to link the pedagogical (including assessment) practices we promote to a compelling portrait of what rhetoric is, why rhetorical arts are important to our society, and what it means to be human and literate, a portrait that clearly demonstrates the necessity of human relationships and interactions in the evaluation of rhetorical abilities. Human writers need human readers, not software. Students need responses from peers and teachers, not computers. The teaching of writing needs to include the assessment of writing, not outsource it.

In defense of these principles and practices we will need to educate ourselves, argue our case to the world, and be ready to fight those who would put their profit before our students' learning.