Speaking My Mind

Robo-Grading and Writing Instruction: Will the Truth Set Us Free?

When I was a graduate student in English literature in the mid-1990s, I wrote a paper on Kenneth Branagh’s (then recent) production of Much Ado about Nothing. My chosen topic was discrepancies between Shakespeare’s text and the seduction scene involving Barachio and Margaret as portrayed in Branagh’s film. When the professor returned the paper to me, I noticed that he had commented on the front page something to the effect that mine was one of the most finely written papers he had received. This comment is not why I remember the paper. His comment on the back of the paper is what has stuck with me all these years. The comment began with a question: “So what?” He went on to posit, quite rightly, that my thesis did nothing to shed light on our understanding of Shakespeare or his play. My essay was, itself, “much ado about nothing” because although well written, it was vacuous and void of substantial critical insight.

I was reminded of this incident last year as I read of a competition sponsored by the William and Flora Hewlett Foundation to see how algorithms could be developed to assess student writing. Hewlett’s education program director, Barbara Chow, concluded, “We had heard the claim that the machine algorithms are as good as human graders, but we wanted to create a neutral and fair platform to assess the various claims of the vendors. It turns out the claims are not hype” (qtd. in Stross). Earlier that year, a University of Akron study found that an analysis of 16,000 middle school and high school essay tests previously graded by humans indicated that, as the researchers stated, robo-graders “achieved virtually identical levels of accuracy, with the software in some cases proving to be more reliable” (qtd. in Bienstock). Dr. Mark Shermis, Akron’s College of Education dean, conceded that “automatic grading doesn’t do well on very creative kinds of writing. But this technology works well for about 95 percent of all the writing that’s out there” (qtd. in Bienstock).

Regardless of the claims of the proponents of robo-graders, the clear winners in the standards movement thus far have been those in the billion-dollar-a-year testing industry. Within the language arts curriculum, the emphasis of the standards movement until now has been focused almost exclusively on reading and its concomitant multiple-choice testing, but as Common Core and other movements place renewed emphasis on writing and critical thinking, the test-prep industry will undoubtedly keep churning out “breakthroughs” in these areas of composition curriculum and assessment.

There are skeptics. Les Perelman, a retired director of writing at the Massachusetts Institute of Technology (MIT), has set out to expose the flaws and foibles of robo-grading. After studying the Educational Testing Service’s e-Rater, Perelman found some of the qualities that e-Rater privileges as “good” writing—sentence structure; length of words, sentences, and paragraphs; and conjunctive adverbs—as evidence of complexity of thought. Where it fails is in recognizing facts, logic, and truth. “E-Rater doesn’t care if you say the War of 1812 started in 1945,” Perelman was quoted as saying in a 2012 New York Times column by Michael Winerip. In fact, Perelman has manufactured essays based on obviously faulty premises and received e-Rater’s top score of six. One such argument reads, “The average
teaching assistant makes six times as much money as college presidents. In addition, they often receive a plethora of extra benefits such as private jets, vacations in the south seas, starring roles in motion pictures” (qtd. in Winerip). While the E.Rater fails to grasp the obviously faulty logic of this argument, the indisputable advantage of the robo-grader over the classroom teacher is speed. E.T.S. researcher David Williamson boasts that E-Rater can grade 16,000 essays in 20 seconds (qtd. in Winerip).

A more recent target of Perelman’s ire is EdX, a nonprofit joint venture of Harvard and MIT to offer online courses promoting essay-grading software as an integral part of its instructional platform. EdX has partnered with other institutions, including Stanford. Grading software is also being used by other “massive open online courses” (MOOCs), including two start-ups founded by Stanford faculty members. Daphne Koller, a computer scientist and founder of one of these MOOCs, Coursera, says of grading software, “It allows students to get immediate feedback on their work, so that learning turns into a game, with students naturally gravitating toward resubmitting the work until they get it right” (qtd. in Markoff). Perelman states that his major concern with the EdX software is that it “did not have any valid statistical test comparing the software directly to human graders” (qtd. in Markoff). Perelman has joined an online petition to stop robo-grading called Professionals Against Machine Grading.

So what?

What’s at Stake in Robo-Grading?

What is most vitally at stake here is not writing assessment or the validity of standardized writing tests. What is most at stake is instruction. In the standards movement, testing drives instruction because the stakes are so high for teachers and schools. When facts, logic, and truth become dispensable in the assessment of writing, then writing instruction, ostensibly, will become focused solely on the mechanics of writing. So much for the short-lived return to critical thinking that the Common Core State Standards initiative promises to bring back to the English curriculum. Although the E-Rater and its brethren may not be interested in the truth, the truth is that writing teachers always have been.

Good writing teachers encourage fact checking, ensure that material is asserted and quoted within its intended context, challenge the logical assumptions of student writing, make personal connections, encourage a student’s development of voice and linguistic ingenuity, and perform a thousand other tasks in responding to student writing that are simply above the pay grade of robo-graders. But will teachers continue to offer rich responses in a high-stakes testing world proliferated with robo-graders, or will they succumb to the pressure to teach to a test that so narrowly defines what “good” writing is?

As a writing teacher for the past 20 years, I have often found myself borrowing my professor’s “So what?” question, using it as a jumping-off point for my responses to student writing whenever a student fails to make a logical connection or, as I did in my Much Ado paper, labors for pages on a banality. As writing teachers know, this type of critical commentary is necessary in teaching students to be critical thinkers. Moreover, “good” writing is always subjective. English teachers are notorious for their pet peeves and personal opinions of what is “good.” Over time, astute student writers will collect these hallmarks of good writing from various teachers, stack them against one another and their own, reject some and embrace others, and eventually develop their own style and criteria for good writing. The adoption of such a narrowly defined concept of writing in which, for instance, each sentence in a student essay must be at least 15 words long or contain a conjunctive adverb, threatens this process. As Patricia O’Connor points out in the May 2012 English Journal, “I know from my teaching experience that the nature of writing is not as linear as the data miners would lead us to believe. Learning how to write well is a complex, recursive process, and despite years of research, it still resists being broken neatly into discreet segments for assessment” (105).
Writing teachers like O’Connor develop what Elliot Eisner refers to as “connoisseurship,” or “the art of appreciation” (215). Of educational connoisseurship, in general, Eisner states, “One must have a great deal of experience with classroom practice to be able to distinguish what is significant about one set of practices or another” (216). Writing teachers develop this connoisseurship of the practices of writing and writing instruction over time through their own reading, writing, and interaction with other readers and writers, as well as through thoughtful reflection on the processes of writing and the teaching of writing. The proponents of robo-grading laud it precisely because it provides some sort of objective quantification of writing, but writing teachers know that a certain degree of subjectivity is inescapable, and indeed even essential to the assessment of writing, as the self cannot be removed from the act of reading (or grading) any more than it can be removed from the act of writing. Students must be taught to read and write in a world where facts matter, where logic is challenged, and where the “truth” is often not only subjective but also subject to nearly inscrutable nuances. In short, they must be taught to write for diverse audiences, not algorithms.

Matthew J. Spireng's book What Focus Is was published in 2011 by Word Press. His book Out of Body won the 2004 Bluestem Poetry Award and was published in 2006 by Bluestem Press. He has several chapbooks including Clear Cut, a signed and numbered limited edition of his poems paired with photographs by Austin Stracke on which the poems are based. Email him at mattspi@aol.com.