

Contents

Teachers vs. Technology—Round One, Fight!	1
The Fight over Computer Grading (Started before Computer Grading)	4
How the Technology Works	7
The Technology’s “Expertise” (Is No More Limited than That of Its Human Counterparts)	9
The Enduring Lure of Labor-Saving Devices	12
(How) Do Labor-Saving Devices Work?	14
Faking It: Why Rich Kids Can Do It, and Poor Kids Can’t	17
The Matter with MOOCs	20
The Problem with “Papers”	23
What They Get Right (and We Get Wrong)	26
Why They’re Still Wrong	29
What We Can Do about It	34
<i>Notes</i>	39
<i>About the Authors</i>	45

The Problem with Education Technology (Hint: It's Not the Technology)

TEACHERS VS. TECHNOLOGY—ROUND ONE, FIGHT!

*In today's society, college is ambiguous. We need it to live, but we also need it to love. . . . Teaching assistants are paid an excessive amount of money. 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, a starring roles in motion pictures. Moreover, in the Dickens novel *Great Expectation*, Pip makes his fortune by being a teaching assistant.¹*

It was our last day at the 2012 Conference of College Composition and Communication (CCCC) in St. Louis, the biggest annual gathering of college writing teachers in the country. We were done with our own presentation, and we could eat only so many ribs in a day . . . so we were looking through the conference program to see what else was on offer . . . when a full-page listing caught our eye. It was for a “Special Session.” Something particularly Big and Important. This Special Session, the program told us, would be a “vigorous but civil debate” on Automated Essay Scoring (AES), the technology that lets computers score writing assignments.

We were intrigued. When an academic conference has to specify that an event will be “civil,” you know something’s up.

The ballroom was vast and long. There were about two hundred seats facing a long table and a giant PowerPoint screen. At one end of the table sat Paul Deane and Chaitanya Ramineini, researchers from Education Technology Services (ETS): the College Board, the minds behind standardized tests like the SAT, GRE, and AP exams, and—along with their for-profit colleagues Pearson and Vantage—the loudest promoters of AES. They were here today to present their newest innovation: the e-Rater[®] Scoring Engine, a computer program that—according to an impressive set of studies—can score a student essay with the reliability of a human.

At the other end of the table sat Les Perelman, the longtime (now emeritus) Director of Undergraduate Writing at MIT—and the best-publicized crusader against AES. He authored the first paragraph of this section, which he fed, along with another page-and-a-half of similar drivel, into e-Rater[®]. And it got a 6, the highest-possible score. The point seemed clear and obvious: machines can’t read. They can’t understand. And they can’t, and certainly shouldn’t, replace humans in educating our children.

(Between them, fittingly, sat Carl Whithaus, director of the University Writing Program at UC Davis, who works to integrate AES into human-centered teaching. We’ll get back to him later.)

Les was the perfect foil to Deane and Ramineini. They were impeccably dressed; he wore a suit that didn’t fit quite right and looked a little sweaty. Their speech was memorized and newscaster-perfect; he spoke off the cuff and made no effort to hide a general-purpose ethnic accent. Their PowerPoints were sleek, branded, and serious; his was homemade and funny, complete with gifs of rabbit holes and the *Twilight Zone*.

None of this was by accident. They were each playing their parts to a T: the bloodlessly efficient technocrats versus the righteous, ruffled, lone defender of flesh-and-blood teaching. “Vigorous but civil” be damned—this was a brawl. An *agon*. A pundit war between good and evil, man and machine. We, the human educators of CCCC, sat and winced as the forces of evil took their best shot—and then rejoiced as our champion tore them limb from limb, down to their last unfounded assumption and logical fallacy.

He was, after all, fighting for our lives. Or at least our livelihoods. Humanistic objections to robo-teaching aside, we were precisely those humans AES stands poised to eradicate. AES, which can grade thousands of student essays in mere seconds, could convince a budget-wary public that small class sizes and individualized instruction are unnecessary luxuries. In a political moment where education funding is under constant attack, it's not hard to imagine administrators and elected officials (even the well-meaning ones) using AES as a rationale to lay us off. So when we see Les up there proving that AES doesn't work, won't work, can't work, of course we cheer loudly.

The problem is, outside the cozy confines of CCCC, he's losing. We're losing. Education technology—computer-based classroom instruction, online K–12 schools, MOOCs (Massive Open Online Courses), as well as AES—is all the rage. It's what's getting all the media attention, foundation funding, and government grants. It's the cutting edge, the thing all forward-thinking principals, superintendents, policymakers, and executives (both public and private) are talking about. Les's dogged defense of old-fashioned teaching might make us cheer, but it's making everyone else yawn.

And maybe most frustrating of all, the science all seems to be on their side. Sure, Les's argument makes intuitive sense—

how could you trust a program that gave that pile of s#t he wrote the highest-possible score?—but then again, how can you argue with scores of reliable, data-driven studies?

This is the problem we'll tackle in this book. We'll explain the technology, how it works, and what it does (and doesn't do). We'll explain where it comes from, and how it's come to take the form that it has, in the classroom and in our culture. We'll also explain why the debate over the technology has taken the shape it has—a shape that stops us from understanding the real problem or doing anything about it. Finally, we'll explain what that real problem is, and how we—teachers, students, citizens—can work to solve it.

Because there is a problem. It's just quite a bit subtler, tougher, and more complicated than the standard “civil debate” would have us believe.

THE FIGHT OVER COMPUTER GRADING (STARTED BEFORE COMPUTER GRADING)

Human versus machine. Good versus evil. Teachers versus technology. Kids versus computers. This is a Big Issue, and not just for writing teachers.

A few weeks after the conference, it made the *New York Times*. “Facing a Robo-Grader? Just Keep Obfuscating Mellifluously,” read the headline. The story, by longtime *Times* reporter Michael Winerip, starts with a recent study that “concluded that computers are capable of scoring essays on standardized tests as well as human beings do”—at the rate of 48,000 essays per minute.

It gives some space to ETS representatives, including Paul Deane. But most of the story belongs to Les, who demolishes the software with his usual gusto:

The e-Rater's biggest problem, he says, is that it can't identify truth. He tells students not to waste time worrying about whether their facts are accurate, since pretty much any fact will do as long as it is incorporated into a well-structured sentence. "e-Rater doesn't care if you say the War of 1812 started in 1945," he said.

Mr. Perelman found that e-Rater prefers long essays. A 716-word essay he wrote that was padded with more than a dozen nonsensical sentences received a top score of 6; a well-argued, well-written essay of 567 words was scored a 5.

An automated reader can count, he said, so it can set parameters for the number of words in a good sentence and the number of sentences in a good paragraph. [. . .] e-Rater likes connectors, like "however," which serve as programming proxies for complex thinking. Moreover, "moreover" is good, too.

Gargantuan words are indemnified because e-Rater interprets them as a sign of lexical complexity. "Whenever possible," Mr. Perelman advises, "use a big word. 'Egregious' is better than 'bad.'"²

The ETS representatives make some important points in response, which we'll get to later. But they don't—and can't—refute Les's basic claims. Computers can't read. They can't understand truth. They can only grade by counting things: length of essays, length of sentences, and length of words. So, obviously, they shouldn't be used in place of humans. Right?

Not so fast. Take a look at this:

In March, Les Perelman attended a national college writing conference and sat in on a panel. [. . .] "It appeared to me that regardless of what a student wrote, the longer the essay, the higher the score," Dr. Perelman said. A man on the panel from the College Board disagreed. "He told me I was jumping to conclusions," Dr. Perelman said. "Because M.I.T. is a place where everything is backed by data, I went to my hotel room, counted the words in those essays and put them in an Excel spreadsheet on my laptop. [. . .]"

He was stunned by how complete the correlation was between length and score. "I have never found a quantifiable predictor in 25 years of grading that was anywhere near as strong as this one," he said. "If you just graded them based on length without ever reading them, you'd be right over 90 percent of the time." The shortest essays, typically 100 words, got the lowest grade of one. The longest, about 400 words, got the top grade of six. In between, there was virtually a direct match between length and grade. He was also struck by all the factual errors in even the top essays. [...]

How to prepare for such an essay? "I would advise writing as long as possible," said Dr. Perelman, "and include lots of facts, even if they're made up." This, of course, is not what he teaches his M.I.T. students. "It's exactly what we don't want to teach our kids," he said.³

This is another story from the *New York Times*. Like the first one, it's about a fight between Les and ETS at CCCC. Like the first one, Les is criticizing ETS for scoring essays based on length, rather than truth or quality. And like the first one, it's pretty darn convincing.

Only this story isn't from March 2012. It's from March 2005. And it wasn't about AES at all.

This story was, rather, about the then-brand-new essay portion of the SAT. (Which is now defunct, due in no small part to Les's efforts.) It seems the SAT essay had all the same problems Les now identifies with AES. But computers didn't grade these essays.

Here's where the plot thickens. Yes there's a problem with the technology, but it's not just about the technology—take away the computers, and the problem persists.

To understand why, first we need to understand more about how the computers work, how they're programmed, and what, exactly, they're programmed to do.