

Software's Essay Test: Should It Be Grading?

By LINDA PERLSTEIN
Washington Post Staff Writer

It took Hugo Rousselin, a sophomore at New Mexico State University, about 20 minutes to tap out his first essay for his psycholinguistics class on the computer in his bedroom. It took less than 20 seconds to find out he had gotten a "B."

Good, the accompanying comments read, but "you need to define the word superiority effect and what it does." So, minutes before class the next day, Rousselin ducked into the computer lab, called up his essay, and did just that. And—seconds later—found out he had raised his grade to an "A."

Rousselin's essay was graded not by his professor, Peter Foltz, but by the computer program Foltz helped design. The technology represents one of the first major efforts in the country that employs computers to evaluate the content of a student's essay, rather than simply check its spelling, grammar

See ESSAYS, A8. Col. 1



The Computer Finally Gets to the Essay Question

ESSAYS, From A1

or adherence to rules of style.

"I think it's kind of creepy to think that a computer is starting to do what only we used to be able to do," Roussein said. On the other hand, "you get immediate feedback. In my other psychology class, I wrote a five-page paper and I didn't get it back until two weeks later."

The technology will be commercially available in a few months, and scores of educators—from elementary schools to universities—have said they are eager to implement the Intelligent Essay Assessor in their classrooms. Even the venerable Educational Testing Service, which administers most of the nation's standardized tests for college and graduate school admissions, hopes by the beginning of next year to be using a similar technology to judge the written essays submitted by takers of at least one of its major exams.

For an outfit that hires several thousand human beings each year to score written responses, computerized grading is an advance that can be implemented "economically and fairly," said Barbara Voltmer, who oversees the Oakland office.

There's no question the technology is efficient. And developers claim it is also reliable. In several studies of the Intelligent Essay Assessor, when two people and the computer graded the same essays, the computer agreed with each grader as often as the graders agreed with each other.

But the mounting interest in the technology has been matched by concern among some educators that, even if such programs can judge the quality of essays the quest for efficiency is coming at the cost of human involvement and could harm students.

Foltz and his partner, Thomas Landauer, who teaches at the University of Colorado in Boulder, say the ideal use of their technology is not necessarily to grade students' work but to give students feedback that might improve their writing in later drafts. "My goal is not to replace teachers. My goal is to have students do more writing," Foltz said.

But he acknowledges there's nothing to stop slothful teachers, eager to unload a burden, from using the computer in place of human involvement.

Landauer began work on the technology 10 years ago, when he was a researcher for Bell Communications, a telecommunications firm.

In essence, the technology is propelled on the idea that, with enough data, a computer can learn to understand the use of language the same way people do. To understand how it works, consider the case of a college professor who wants to assess a couple hundred students taking his Psychology 101 class. The professor feeds into the computer the Introduction to Psychology textbook. Once the computer has mathematically analyzed the book's language, Landauer said, it will comprehend the words in much the same way people do: What they mean, how they are used, how different concepts relate and what would be appropriate ways of describing them.

To grade a specific assignment, the professor can do one of two things. He can feed the computer a "gold standard" essay and tell the computer that anything close to it receives the highest grade. Those that fall below that ideal rank according to how far off the mark they land. Alternatively, the professor can feed into the computer a set of about 20 essays that have been graded, so that the computer knows what constitutes an A paper, a B paper, and so on. Then the computer reads the students' submissions, one by one.

Test Results

A version of the Intelligent Essay Assessor—which grades middle school students' efforts on summarizing a text, but not their spelling or grammar—gave both of these compositions 98 out of 100 points and the comment "Great! I think your summary is awesome." Though the essay at top is clearly inferior, Thomas Landauer, who created the grader, said a sixth-grader couldn't have produced it without grasping the most important concepts.

Solar Energy
By Linda Perlstein

The sun produces lots of energy every day. You can make that energy into useful for the house.

Catch it in a solar collector. The sun absorbed and made to heat. The dark metal plates because black is the best color and passes inside glass boxes. Water can be heated in tanks.

Methods like photovoltaics. Sunlight gets a solar cell and silicon, and electrons dance within creates a current. Or solar thermal systems, sunlight passes through a mirror and heats liquid. Steam drives a generator.

Photovoltaics is too expensive. Thermal plants like LUV shut down because of falling oil prices and over-budget construction project.

Solar power has great potential for making energy for the earth.

Solar Energy
By Linda Perlstein

Every day, the sun produces more than enough energy to fuel the earth. Capturing that energy and transforming it into heat and electricity is tricky, since it is so spread out and also varies depending on the weather, time of day and season.

One way to capture solar energy is a solar collector. As the sun passes through dark metal plates inside glass boxes, it is absorbed and changed into heat, and then distributed throughout the house. It can also be sent through tanks, where water is heated.

Similar, but more complex, methods are used to transform solar energy into electricity: with photovoltaics, sunlight strikes a solar cell, made up of silicon, and the movement of the electrons within creates a current. With solar thermal systems, sunlight passes through a mirror and heats liquid, which makes steam, that drives a generator.

Sunlight is unlimited, and therefore solar energy has great potential to energize the earth.

THE WASHINGTON POST

For each essay, the computer looks at the combination of words used and assesses the content and meaning, and compares it to what it has already seen. It then assigns a score, and reports how confident it is in that score.

Because the program already has a "vocabulary" it learned from the textbook, Landauer said, it knows which words and phrases mean the same thing and can compare essays that don't use the same terms. In this way, it goes beyond less-sophisticated attempts of computerized grading systems, which only looked to words that matched exactly.

If the computer sees the word "doctor" used with "patient" and "surgeon" used in similar contexts with "patient," whether or not it's seen "doctor" with "surgeon," the computer assumes they are related. It recognizes that "the doctor operated on a patient" is similar to "the surgeon used his scalpel."

Landauer and Foltz have been testing the Essay Assessor on their own students and have created a

company through which they hope to market their product to both institutions and individual teachers. They haven't decided yet how much to charge. "But I can say," Foltz said, "it's less than going through the essays by hand."

The idea of computerized essay grading appeals to many educators, who generally agree that while they loathe the seemingly Sisyphean task of grading them, compositions are far superior to multiple-choice exams as a learning and testing tool. Several government and other agencies that have to certify thousands of teachers or accountants each year, have also expressed interest.

Phylis Floyd, who teaches art history at Michigan State University, said she would not employ the software as a surrogate grader of student essays, but she would find it useful for training teaching assistants how to

grade. "There are too many of us who are swayed by the slick, glossy use of language and overlook that someone didn't pay attention to the details," she said. "It can help to objectify what is really a subjective process."

Florida State University has also expressed interest in the technology, and if the results of planned experiments are satisfying, its School of Information Studies—the modern version of library science—plans to use the software to score the 200 finals in one of its introductory courses. Professor Myke Gluck is intrigued not only by the time that could be saved but also by the potential consistency a computer could deliver. "I believe human beings are more accurate," he said, "but the damn machine's reliable as hell."

If the technology is gaining supporters, however, it is also drawing concerns from educators who think computerized essay graders are more an insidious omen of a coolly technological society than a valuable teaching device. Anne Greene, an English professor at Wesleyan University in Connecticut and director of the writers conference there, finds the notion "depressing." To Greene, writing is not about including the ideal pieces of information but about weaving thoughtful arguments into compelling prose. "This is why so many people write poorly," she said. "You put the pieces in the right place and you get a 100."

Indeed, even its creators say their essay grader isn't appropriate for creative writing assignments, and if it's used for that, it would be a mistake. When a fact-based assignment is answered creatively—for example, if a student uses a sports analogy with words the computer is not familiar with—the Essay Assessor fumbles, and tells the professor to grade it by hand.

"The computer can't look at it like a human being can look at it and say, 'Does it make sense?'," said Ron Lamb, a sixth-grade teacher at Platt Middle School in Boulder, where the technology is being tested. "It just compares it to the original text."

But the machine is useful, he and colleague Cindy Matthews say, for helping students identify when they've fallen into what they see as the two worst habits of sixth-grade writers: leaving out whole topics, or rambling on.

Educational Testing Service says its program, the E-Rater, is being considered for, among others, the business school entrance exam (GMAT) and the Graduate Record Exam, and could eventually be used on the ubiquitous Scholastic Assessment Test, once that test is taken on computer. ETS says that its technology goes a step further than the Essay Assessor by using linguistic analysis that can judge whether an essay has a desirable argumentative structure, not just key concepts. But, developer Karen Kukich concedes, "it doesn't do nuances."

Nor does it socialize. Voltmer, who spent 12 years overseeing West Coast essay scoring for ETS, already speaks nostalgically of sitting in hotel conference rooms with hundreds of other educators who earn extra cash each year by grading exams. It's a reunion of sorts, where they can shoot the breeze, share their ideas about teaching, talk about books. She sighed at the recollection, then the administrator in her said, "Hiring human beings to score the tests is a very expensive proposition."