

Software Said To Grade Essays for Content

By Debra Viadero and Steven Drummond

San Diego

Researchers at the University of Colorado at Boulder have developed a computer-software program that they say can grade student essays for content just as well as teachers can.

"We're not seeing this as a replacement for teachers," said Peter W. Foltz, one of three developers of the program, known as Intelligent Essay Assessor. "But we think it may help teachers have more time to interact with students."

The program was created and tested over a period of 10 years by psychologists at the university, who initially sought to create a computer model that would mimic the way people learn about words and language. But, along the way, they said, they discovered that their theoretical model could also be used for more practical purposes, such as grading essays or giving students diagnostic practice in summarizing material.

Since then, the researchers have tested the program with almost 1,200 essays.

"From 6th graders to first-year medical students, we get consistently good results," said Thomas K. Landauer, a University of Colorado psychology professor and the project's lead investigator. "The agreement between a human and the machine is about exactly the same as between two humans."

But unlike humans, the computer will always give the same grade to the same essay, the researchers said.

"It doesn't matter what else it was doing that day or if it was tired," said Darrell Laham, a University of Colorado doctoral student who's also taking part in the project. "It doesn't ever have a bad day."

The findings were presented here last week at the annual meeting of the American Educational Research Association.

Researchers have experimented for years on ways to use computers for grading essays. Ellis B. Page, a Duke University professor of psychology and education, has developed a program called Project Essay Grade, or peg, which assesses the quality of writing in students' essays. It has shown similarly good results. ("**Making the Grade**," May 31, 1995.)

But the program unveiled last week may be among the first to grade essays for content.

The Colorado researchers have applied for a patent for the program, but say they may be a few years away from marketing it commercially. The software requires a computer with about 20 times the memory of an ordinary personal computer.

The researchers predict that in five years, however, the program will be able to run on most computers.

More Time for Teachers?

Ultimately, researchers say, such software could be a boon for professors and teachers as well as for commercial testing companies that grade thousands of exams.

"Teachers in middle schools report that they spend six-and-a-half hours a week grading essays, and high school teachers spend more than ten-and-a-half hours," Gary B. Struck, a professor of educational psychology at the University of North Carolina at Chapel Hill, said during a conference session on the PEG project.

In addition, many teachers lack confidence in the grades they give on essays and would like some means of assessing their scoring, he said. "They'd like to have some kind of comparison."

The developers of Intelligent Essay Assessor said their program might also be a useful supplement to textbooks published by commercial textbook companies. Students could use the program to test their knowledge of the textbook material.

Hard To Beat

The new software program grades essays by evaluating how closely the students' words approximate content already fed into the program through textbooks and model essays that professors have already graded by hand.

"You could write something like, 'The doctor operated on the patient,' and if the text said, 'The physician did surgery,' the computer would see those things as very similar," said Mr. Foltz, who is also an assistant professor of psychology at New Mexico State University.

Longer essays also get better grades because researchers have found that essay length correlates closely with the quality of the content.

Critics of the program have suggested that students could easily fool the computer by including the right content but putting it in the wrong order.

Students could still get a good grade that way, the researchers said, "but it's hard to say good ideas without saying it well," Mr. Foltz said.

What's more, the computer flags essays that are "too creative" or that contain a lot of inconsistencies. It brings such essays to the attention of human graders. Plagiarized essays also get flagged.

The researchers have spent a lot of time trying to fool the system, Mr. Landauer said, but "the easiest way to cheat this system is to study hard, know the material, and write a good essay."