

Cheaters Find an Adversary in Technology (Dec 10)

By TRIP GABRIEL

Published: December 27, 2010

Mississippi had a problem born of the age of soaring student testing and digital technology. High school students taking the state's end-of-year exams were using cell phones to text one another the answers.



Drew Angerer/The New York Times

John Fremer, 71, a Caveon co-founder who was once the chief test developer for the SAT.

A High-Tech Approach

With more than 100,000 students tested, proctors could not watch everyone — not when some teenagers can text with their phones in their pockets.

So the state called in a company that turns technology against the cheats: it analyzes answer sheets by computer and flags those with so many of the same questions wrong or right that the chances of random agreement are astronomically small. Copying is the almost certain explanation.

Since the company, [Caveon Test Security](#), began working for Mississippi in 2006, cheating has declined about 70 percent, said James Mason, director of the State Department of Education's [Office of Student Assessment](#). "People know that if you cheat there is an extremely high chance you're going to get caught," Mr. Mason said.

As tests are increasingly important in education — used to determine graduation, graduate school admission and, the latest, merit pay and tenure for teachers — business has been good for Caveon, a company that uses "data [forensics](#)" to catch cheats, billing itself as the only independent test security outfit in the country.

Its clients have included the [College Board](#), the Law School Admission Council and more than a dozen states and big city school districts, among them Florida, Texas, Washington, D.C., and Atlanta — usually when they have been embarrassed by a scandal.

"Every single year I've been in testing there has been more cheating than the year before," said John Fremer, 71, a Caveon co-founder who was once the chief test developer for the SAT.

Exposing cheats using statistical anomalies is more than a century old. James Michael Curley, the so-called rascal king of Massachusetts politics, and an associate were shown to have copied each other's civil service exams in 1902 because they had 12 identical wrong answers.

Probability science has come a long way since then, and Caveon says its analysis of answer sheets is the most sophisticated to date. In addition to looking for copying, its computers, which occupy an office in American Fork, Utah, and can crunch up to one million records, hunt for illogical patterns, like test-takers who did better on harder questions than easy ones. That can be a sign of advance knowledge of part of a test.

The computers also look for unusually large score gains from a previous test by a student or class. They also count the number of erasures on answer sheets, which in some cases can be evidence that teachers or administrators tampered with a test.

When the anomalies are highly unlikely — their random occurrence, for example, is less than one in one million — Caveon flags the tests for further investigation by school administrators.

Although its data forensics are esoteric and the company operates in the often-secretive world of testing, Caveon's methods are not without critics. [Walter M. Haney](#), a professor of education research and measurement at [Boston College](#), said that because the company's methods for analyzing data had not been published in scholarly literature, they were suspect.

"You just don't know the accuracy of the methods and the extent they may yield false positives or false negatives," said Dr. Haney, who in the 1990s pushed the [Educational Testing Service](#), the developer of the SAT, to submit its own formulas for identifying cheats to an external review board.

David Foster, the chief executive of Caveon, said the company had not published its methods because it was too busy serving clients. But the company's chief statistician is available to explain Caveon's algorithms to any client who is curious.

Other means that the company uses to stop cheating are not based on statistics.

For the Law School Admission Council, which administers the LSAT four times a year to a total of more than 140,000 people, Caveon patrols the Internet looking for leaked questions on sites it calls "brain dumps," where students who have just taken an exam discuss it openly.

"There's all kinds of stuff on the blogs after the test trying to guess which stuff will show up in the future; there's a whole cottage industry," said Wendy Margolis, a spokeswoman for the council.

Caveon, which declined to reveal what it charges clients, sends letters to the people who operate those Web sites requiring them to take down the material under the [Digital Millennium Copyright Act](#).

Standardized testing is controversial with some parents and educators, but not to Dr. Fremer, Caveon's longtime president, who recently gave up managerial duties. He credits testing with helping him escape from a working-class background. The son of a New York City firefighter, he earned a Ph.D. from Columbia in educational psychology and measurement, and then went to work for the Educational Testing Service. He first worked in the verbal aptitude department and later spent seven years leading a major overhaul of the SAT in 1994.

Dr. Fremer has little patience with critics who say standardized tests do not accurately measure academic prowess.

"Fundamentally," he said, "testing is a way of ascertaining what you know and don't know and developing ranks, and the critics go right to the ranks. Well, it does rank, but on the basis of knowledge of the subject, and if you think that's not important, there's something improper about the way you think."

More ruffled academic than business type, Dr. Fremer has an air of great confidence and interest in his own ideas. He likes to tell stories, which frequently devolve into lengthy digressions. His home office near Atlantic City is the lair of an eccentric, packed with collections of casino matchbooks (he does not smoke) and empty cigar boxes (he thinks about turning them into pocketbooks).

"At this stage of my life, I'm an icon," he said without an iota of self-consciousness.

Although it is in Caveon's interest to dramatize or even inflate the incidence of cheating, the company was criticized this year by a state governor for underestimating it.

Hired to analyze English and math tests from Atlanta students after a state audit identified dozens of schools where cheating might have occurred, Caveon found far fewer problems. It identified a dozen

elementary and middle schools at which cheating had probably taken place, but it essentially exonerated 33 others on the state's list of suspect schools.

Gov. [Sonny Perdue](#) criticized that conclusion and appointed his own investigators in August. In an [interview with The Atlanta Journal-Constitution](#), he accused Caveon of seeking to “confine and constrain the damage” and suggested it was trying to protect its business prospects with other school districts.

Dr. Fremer dismissed that suggestion. Caveon's data forensics on answer sheets were more sophisticated, he said.

The state had looked at just one metric: the number of times wrong answers had been erased and changed to right ones. The schools it identified as suspect had a statistically higher rate of wrong-to-right erasures than the statewide average. It inferred that adults had tampered with the tests.

Caveon maintains that counting wrong-to-right erasures is only one of several ways to mine answer-sheet data, and it can lead to false accusations. Dr. Fremer said it was common, for example, for students to lose their place in a test and erase a string of answers once they realized the mistake.

“Our analysis was better,” he said. “It was more in-depth. It didn't inflate small differences and make a lot out of them.”

Caveon's philosophy is that it is not necessary to ensnare every cheat to reduce cheating over all. Since cheats rarely confess even when confronted with overwhelming evidence, it is better to identify the most egregious cases and ignore the borderline ones.

“Your goal is not to catch a bunch of people and hang them,” Dr. Fremer said. “Your goal is to have fair and valid testing.”

“Prevention is the goal,” he said, as matter-of-fact as Joe Friday. “Detection is a step. We detect and prevent.”

This article has been revised to reflect the following correction:

Correction: December 31, 2010

An article on Monday about Caveon Test Security, a company that analyzes answer sheets from standardized tests to identify cheating, misstated an example of a statistical threshold the company would use to identify suspicious tests. Tests with anomalous answers would be flagged as suspicious if the chance of the anomalies occurring randomly is less than one in one million, not greater than one in one million.

©2010 New York Times