



How Test Companies

FAIL YOUR KIDS

Shane Fulton of Yardley, Pennsylvania, suffered anxiety and sleepless nights after a 390-point error dropped his SAT score.

By David Glovin and David Evans

◆ Jerry Lee Faine Elementary School in Dothan, Alabama, starts each day with two hours of reading and vocabulary. After that, there's arithmetic. "If you can read, you can do anything," says Principal Deloris Potter, a spry woman of 59 who has run the school since 2002.

Potter, trusting the work of her teachers, was confident of passing grades in April 2005 as students began two weeks of mandatory standardized testing in reading and math. That July, state education officials told Potter her school had failed the Alabama Reading and Mathematics Test. The state warned it might fire teachers if scores didn't improve, she says. A dozen students transferred after the substandard rating. Faculty morale plunged. "We felt like dogs," says Charlotte Adams, a reading specialist at the school.

In February 2006, the state said Jerry Lee Faine Elementary had passed. Harcourt Assessment Inc., a unit of London-based Reed Elsevier Plc and one of the world's largest test companies, had improperly graded the exam. The snafu is at least the 30th time since 2000 that San Antonio, Texas-based Harcourt Assessment, which also wrote the exam, has made errors such as improper scoring, faulty instructions and questions with more than one answer.

Harcourt isn't alone. Other companies are constructing flawed tests, administering them improperly and scoring them incorrectly, according to lawsuits and education department records in 15 states. In March, Pearson Assessments, a unit of London-based Pearson Plc, the world's biggest educational publisher, had to explain to high schoolers across the U.S. that it had erred in scoring about 5,000 SAT college entrance exams because its scanners couldn't read answer sheets that had expanded from humidity. The next month, education officials in Minnesota discovered a separate issue with answer sheets that Pearson Assessments had created for a state-mandated exam. At least 500,000 people taking tests from 2000 through '06—from Nevada third graders to aspiring teachers in many states—were victims of test company mistakes, documents show.

"The errors we've seen from testing companies are probably just the tip of the iceberg," says David Berliner, 68, Regents' Professor of Education at Arizona State University in Tempe, who has written more than 200 articles, books and book chapters about education and

The \$2.8 billion industry hires \$10-an-hour graders for exams that control U.S. schools. Prepping students brings 20 percent profit margins.

served as president of the 25,000-member American Educational Research Association. "State education departments often lack the ability to adequately supervise these companies."

The U.S. is in a testing frenzy. Students in the 92,816 American public schools will take at least 45 million standardized reading and math exams this year. That will jump to 56 million in the 2007-08 school year, when states begin testing science as part of the 2002 federal No Child Left Behind law, the most comprehensive education overhaul in half a century. Beyond No Child, tens of millions of additional tests assess college hopefuls, certify future stockbrokers and even evaluate preschoolers. With the stakes for making the grade so high for so many, errors by test companies have dramatic consequences. Joseph Conigliaro lost his Pennsylvania teaching job after Princeton, New Jersey-based Educational Testing Service, the world's biggest standardized test company, incorrectly scored three of his licensing exams. ETS, which will pay \$11.1 million to 4,100 teachers who were falsely failed, called the error an "anomaly." Ryan Beck & Co. asked Linda Cutler to resign from a senior associate job at the securities firm after she and 1,881 other test takers were scored incorrectly last year on the Series 7 licensing exam for securities representatives. (See "How NASD Flunked a Pro," page 134.)

"It's an exponentially growing catastrophe," says James Popham, an emeritus professor of education at University of California, Los Angeles, and author of 25 books on education. "No one knows how bad it is, and it's going to get worse."

Deputy U.S. Education Secretary Raymond Simon says states must better oversee test companies. "The whole teaching system is based on the results of those tests," Simon, 61, says. "If the integrity of the testing process is called into question, that brings into question the whole accountability system."

The national obsession with performance and measurement means a booming business for test-producing and grading companies. In 2005, CTB/McGraw-Hill, Educational Testing Service, Harcourt Assessment, Pearson Assessments and smaller firms generated \$2.8 billion in revenue from testing and test preparation, according to Boston-based research firm Eduventures LLC. No Child tests alone produced about \$500 million in annual revenue in 2005-06.

Along with creating exams, Harcourt Assessment, Pearson Assessments and companies such as White Plains, New York-based Houghton Mifflin Harcourt sell mass-produced workbooks, practice tests and computer software that teachers use year-round to prepare students for No Child and other tests. The burgeoning test preparation industry generated \$1.7 billion in annual revenue last year. The \$1.1 billion testing market and the \$1.7 billion test preparation business will grow by a combined 30 percent by the 2009-10 school year, Eduventures predicts.

For test companies, pitching schools to buy preparation materials after receiving a No Child contract is routine, says Robert Schaeffer, public education director at the National Center for Fair & Open Testing, a Cambridge, Massa-

chusetts-based nonprofit group. "It's standard business practice, the equivalent of razor companies' giving away razors so they can make money selling blades," he says. "It's where the real profits are."

Profit margins in test preparation are as much as seven times higher than they are for No Child tests, partly because there are no requirements for high-quality questions on practice exams. States leave it to schools and school districts to decide whether the test preparation materials they're buying are sound. Houghton Mifflin Harcourt, publisher of the Buckle Down test preparation workbooks, reported operating margins of 21 percent in its test preparation division for the first half of 2006. (See "Test Prep: A+ in Profits, Incomplete in Results," page 136.)

In comparison, No Child tests, which must be custom designed for almost every state, have pretax profit margins as low as 3 percent, says Kurt Landgraf, chief executive officer of Educational Testing Service. He says his not-for-profit company lost \$2.6 million on a \$236 million four-year No Child contract in California.

Richard Rizzo, chief financial officer of Measured Progress Inc., a Dover, New Hampshire-based nonprofit firm that produces No Child tests, says he expects to earn margins triple those of No Child exams by selling practice questions and tests that schools use to gear up for the actual exams. Getting a foot in the door with a No Child contract can also lead to sales of achievement or psychological tests not related to No Child. "Companies could conceivably low-ball the customized test because they know they could go in and sell the off-the-shelf products with a 40-50 percent margin," says Rizzo, 62, referring to tests that aren't specially designed for individual states.

Whether or not they low-ball, companies often scrimp when they bid on No Child contracts, Eduventures analyst Tim Wiley says. Getting a contract involves the same process as selling supplies or cafeteria food to a school: A company submits what it expects to be a winning package. "As with any bidding situation, it definitely requires a lot of cost cutting," Wiley says. "Or, in some cases, cutting corners."

In Florida, CTB/McGraw-Hill won part of the state's testing contract for 3,800 schools in 2005. To grade the essay portion, the Monterey, California-based unit of McGraw-Hill Cos. hired \$10-an-hour workers from Kelly Services Inc., the second-largest U.S. provider of temporary employees, and other companies. Among the 2,947 graders was a person who won the job while he was employed packing bags of potato chips for PepsiCo Inc.'s Frito-Lay unit, applications compiled by the Florida state senate show. Kelly spokeswoman Renee Walker declined to comment.

Another grader was a cook in an Orlando, Florida, diner. One essay evaluator wrote he was "laid off" from a clerical job after working as a janitor. He graduated from Ambassador University, a Worldwide Church of God-run school in Big Sandy, Texas, in 1997. The school shut down that same year. Another said that he majored in "Phylosophy/Humanity" at Mount Angel Seminary in St. Benedict, Oregon.

Steven Weiss, vice president for communications at Mc-

Graw-Hill, said in an e-mailed statement that the company had performed extremely well in scoring more than 90 million documents with a total of more than 755 million essay and short-answer questions during the past five years.

CTB/McGraw-Hill, Harcourt Assessment and Pearson Assessments don't break down their revenue from No Child tests and preparation materials in regulatory filings. Public records from the Wyoming department of education show the state is paying Harcourt, which has a \$21 million, four-year No Child contract, more than \$120 per student each year. Of that, about half is for No Child tests, and the rest is for preparation materials and other testing products.

Schools in Okaloosa County, Florida, pay \$9.50 per student for a series of preparation tests called Stanford Learning First, which Harcourt Assessment renamed Learnia. By comparison, Harcourt received \$4.93 per child from the state of Florida in 2005 to develop questions for its No Child-mandated annual Comprehensive Assessment Test.

Harcourt Assessment's experience shows how winning a No Child bid can be a prelude to more sales. In 2004, Harcourt got a four-year, \$44.5 million contract to develop and score Illinois's No Child exams. Chicago schools then began purchasing Harcourt materials, testing director Xavier Botana says. The preparation products included Stanford Learning First practice tests that measured student progress as they prepared for No Child exams. In the 2005-06 school year, the district spent \$1.8 million on Harcourt's new Stanford Learning First product.

Christine Rowland, a former teacher of English as a second language who now trains colleagues at Christopher Columbus High School in the Bronx, New York, says her pupils didn't learn more because of increased testing. Still, she relied on test preparation materials to help students pass the math test. The cost of failure was too high, she says. "If I know they are going to test six things six weeks from now, that's what I'm going to teach," Rowland, 46, says. "It puts a tremendous amount of pressure on. The real fear is that it turns students off from learning."

Test companies, aware that Rowland and other teachers are being judged by how students do on No Child exams, are

of teacher training at Public School 48 in the Bronx.

At Cracker Trail Elementary School in Sebring, Florida, 11-year-old Alexis Szoka took dozens of practice exams last year leading up to the Florida Comprehensive Assessment Test. She wound up a nervous wreck. "My daughter has such test anxiety, she can't take a test anymore," says Alexis's mother, Carol Szoka.

One exam measured whether Alexis understood vocabulary and another checked her spelling. The school tested how well she read and whether she knew math. Some tests compared her reading and math skills with those of other fourth graders. Alexis was evaluated on phonics, writing and her understanding of text on a computer. Most tests were given two, three or four times a year. Teachers gave chapter tests in reading and math and benchmark tests throughout the year to see whether Alexis was progressing.

Andrew Lethbridge, Cracker Trail's vice principal, says one test gave fourth graders practice in filling in answer sheet bubbles on other tests. The materials came from divisions of Harcourt Assessment, Pearson Assessments and smaller, privately held companies.

"It was never like this," says Carol Szoka, who has two grown children who went through the same schools in Sebring, which is 85 miles (137 kilometers) south of Orlando. "They had an achievement test. They just took it. They weren't prepped."

Richard Demeri, Cracker Trail's principal, says test preparation materials have helped his students. Seven years ago, the school was given a grade of C by the state. Now, with test scores higher, the school has an A from the state and is no longer on probation. "There's very little spray-and-pray teaching going on—where you spray everybody and pray they get it," he says. The school uses test results to analyze each student's progress. "It's much more individualized now," he says.

Even if Demeri's students are prepared to take No Child tests, two Florida state senators question whether CTB/McGraw-Hill has qualified people to grade them. Senators Walter "Skip" Campbell and Leslie "Les" Miller Jr. sued the state education department and CTB/McGraw-Hill earlier this year to obtain applications of test graders. The department had refused to release the applications, citing confidentiality. CTB/McGraw-Hill settled the suit by providing copies of the scorers' personnel files with personal identifying information removed.

CTB/McGraw-Hill's \$82 million, three-year Florida contract requires a scorer to have a bachelor's degree in mathematics, reading, science, education or a related field. On its Web site, the Florida Department of Education assures parents that graders of the Florida Comprehensive Assessment Test are professional, trained scorers.

Information the senate obtained shows one grader had an associate's degree, which is below a bachelor's, from the University of Delhi's School of Correspondence Courses and Continuing Education in Delhi, India. She worked as a \$7.50-

One essay evaluator wrote he was 'laid off' from a clerical job after working as a janitor.

inundating schools with ads for preparation products such as practice tests, software and banks of sample questions. Often they say their materials are designed specifically to help students pass the state's No Child test. "I'm getting mail from companies I've never heard of," says Susan Friedwald, head

One Child Left Behind

Tyler Stoken was a well-behaved fourth grader who enjoyed school, earned A's and B's and performed well on standardized tests. In May 2005, he'd completed five of the six days of the Washington State Assessment of Student Learning exam, called WASL, part of the state's No Child Left Behind test.

Then Tyler came upon this question: "While looking out the window one day at school, you notice the principal flying in the air. In several paragraphs, write a story telling what happens."

The 9-year-old was afraid to answer the question about his principal, Olivia McCarthy. "I didn't want to make fun of her," he says, explaining he was taught to write the first thing that entered his mind on the state writing test. In this case, Tyler's initial thoughts would have been embarrassing and mean. So even after repeated requests by school personnel, and ultimately the principal herself, Tyler left the answer space blank. "He didn't want them to know what he was thinking, that she was a witch on a broomstick," says Tyler's mother, Amanda Wolfe, sitting next to her son in the family's ranch home three blocks from Central Park Elementary School in Aberdeen, Washington.

Because Tyler didn't answer the question, McCarthy suspended him for five days. He recalls the principal reprimanding him by saying his test score could bring down the entire school's performance. "Good job, bud, you've ruined it for everyone in the school, the teachers and the

school," Tyler says McCarthy told him.

Aberdeen School District Superintendent Martin Kay ordered an investigation. "My suspension was for refusal to comply with a reasonable request, and to teach Tyler that that could harm him in the future," McCarthy told an investigator. "I never, for a second, questioned my actions."

Tyler, who's 4 feet (1.2 meters) tall and weighs 70 pounds (32 kilograms), hasn't been the same since, his mother says. "He liked the principal before this," she says. "He cried. He didn't understand why she'd done this to him."

Now, Tyler blows up at the drop of a hat, his mother says. "They created a monster. He'll never take that test again, even if I have to take him to another state," she says.

Tyler's attitude about school changed. He became shyer. He's afraid of all tests and doesn't do as well in classes anymore, his mother says.

McCarthy's May 6, 2005, letter to Tyler's mother detailed her son's suspension. "The fact that Tyler chose to simply refuse to work on the WASL after many reasonable requests is none other than blatant defiance and insubordination," McCarthy wrote. In the letter, she accused Tyler of bringing down the average score of the other 10 students in his class. "As we have worked so hard this year to improve our writing skills, this is a particularly egregious wound,"

McCarthy wrote.

Her accusation was wrong, state regulations show. There is no averaging of the writing scores. Each student either meets or fails the state standard.

Tita Mallory, director of curriculum and assessment for the Aberdeen School District, says school officials feel tremendous pressure because of the high-stakes tests. While there's no academic effect on elementary school children taking the exams, there can be repercussions for school administrators. When schools repeatedly fail to show adequate yearly progress, as defined by No Child, the principal can be fired.

"In many ways, there's too much emphasis on the test," Mallory says. "I don't want that kind of pressure on our kids." Out of 74,184 fourth graders taking the WASL test last year, 42.3 percent failed to meet the state standard for writing.

Juanita Doyon, director of Mothers Against WASL and author of *Not With Our Kids You Don't! Ten Strategies to Save Our Schools* (Heinemann, 144 pages, \$14.95), says Tyler's experience is representative of what's wrong with tests like the WASL. "They took a student who loved his school and crushed his spirit," Doyon, 46, says. "We've elevated test scores to be the most important part of school. The principal and teachers are so pressured by the test that they've lost good sense in dealing with children."

DAVID EVANS

an-hour cashier at a duty-free shop at O'Hare International Airport in Chicago before being hired to grade exams, according to the settlement documents. CTB/McGraw-Hill now says this person never scored exams. A personal trainer with a degree in sports science from the University of Leipzig in Germany also graded essays, as did a convicted shoplifter who graduated from West Virginia University with a degree in physical education, the applications show. A person from Hungary wrote he was a "pyshical education" major. A physical education major from Methodist College in Fayetteville, North Carolina, wrote that she had attended "Methidist College."

McGraw-Hill's Weiss said its scorers from the University of Delhi met the requirements for a bachelor's degree. "Individuals must undergo a comprehensive training process before becoming qualified to score," Weiss wrote. "Scorers must maintain performance quality throughout the process."

CTB/McGraw-Hill spokeswoman Kelley Carpenter says the company subjects scorers to a rigorous three- to five-day training program. Next year, at Florida's request, the company will ensure that scorers have appropriate backgrounds for the subjects they grade, she says. "They are constantly monitored," she says. "And if they don't match the quality performance standards, they're not retained as scorers."

Carpenter says spelling errors on an application don't disqualify someone from being hired as a scorer. "Spelling in and of itself is not a requirement," she says.

When Deputy Education Secretary Simon is shown misspellings on applications of Florida scorers, he says he would demand excellence. "It's absolutely important that the integrity of the scorers is something the companies would be proud of and feel comfortable with," he says. "I can't imagine they would feel comfortable with a nonspeller."

Cornelia Orr, head of the Florida Office of Assessment and Performance, says she reviewed about 25 percent of the grader applications. "I felt like CTB had minimally met our expectations," she says. "I know there are ways they can improve."

One reason for the testing foul-ups and their dire effects is

'Spelling in and of itself is not a requirement,' a CTB/McGraw-Hill spokeswoman says of exam graders.

that there's no federal oversight of the testing industry. When the U.S. Congress authorized the No Child law it didn't create an agency to evaluate whether the companies making and selling the exams do an adequate job. Each state oversees its own test contractor.

Roderick Paige, who ran the No Child program as U.S. education secretary from 2001 to '04, says the law is a good one.

Testers' mistakes

Hundreds of thousands of students and some schools got penalized or got the wrong grades because of standardized-testing errors.

MISTAKE	REASON GIVEN
EDUCATIONAL TESTING SERVICE fails 4,100 aspiring teachers on the national Praxis certification test.	<i>Company blames a "statistical anomaly" during scoring.</i>
HARCOURT error leads to four schools mistakenly failing and 10 mistakenly passing Alabama test.	<i>Company uses wrong data in computer scoring program.</i>
MEASUREMENT INC. tells 890 high school students they failed the Ohio Graduation Test.	<i>Company improperly converts raw scores to scaled scores.</i>
PEARSON gives 5,000 students incorrect scores on nationwide SAT.	<i>Company scanners can't read score sheets.</i>

Mistakes were made or discovered from 2003 to '06. Sources: Company press releases, lawsuits, state education department documents

He says his concern is that testing may not be done accurately and competently. Paige, 73, says he summoned top executives from 20 testing companies to a conference room at the U.S. Department of Education on Feb. 20, 2003, and demanded better performance. In 2005, the Education Department's inspector general announced plans to study whether there's a need for federal review to detect and prevent errors. The study isn't yet under way, spokeswoman Catherine Grant says.

"We've got to get better testing producers," says Paige, who's now chairman of Chartwell Education Group LLC, a Washington-based school consulting company. "They're making mistakes."

Harcourt Assessment is making the most errors, according to records in 15 state education departments. In addition to erroneously failing Jerry Lee Faine Elementary, Harcourt wrongly flunked three other Alabama schools because of its grading snafu. It mistakenly passed 10 Alabama schools that should have failed, the state said.

In Connecticut, Harcourt Assessment reported the wrong reading test scores for 355 high school students in 51 districts last year. The state fined the company \$80,000. In Georgia, Hawaii, Illinois, Massachusetts and Virginia, Harcourt made errors on No Child tests and achievement tests given to measure how students compared with one another. States fined the company hundreds of thousands of dollars.

"Employees took shortcuts," Harcourt Assessment Senior

Vice President Robin Gunn wrote in a May 28, 2004, letter to Hawaii school principals, promising stricter oversight. Gunn has since left the company. Hawaii hired a not-for-profit firm, Washington-based American Institutes for Research, to develop and score the tests after discovering more errors on Harcourt's 2005 exams. Illinois also replaced Harcourt in the middle of its contract; Connecticut, Massachusetts and Virginia didn't renew their contracts with the company.

Nevada fired Harcourt in 2004, after the company mistakenly failed hundreds of students, gave inflated scores to thousands of others and produced tests with missing pages, misspellings and flawed instructions, according to Nevada Education Department records.

"It was errors, one after the other, and not to a single student but to a large number," says Karlene Lee, the assistant superintendent in Clark County, Nevada, which includes Las Vegas. "In education, we don't have the luxury to say that

2 percent doesn't matter. Every child has to be accurate."

Nevada fined Harcourt Assessment \$425,000 in 2002, before firing the company. Harcourt's approximately \$290 million in revenue last year was 3 percent of Reed Elsevier's sales, according to company filings. Reed Elsevier reported its profit increased 62 percent in the six months ended on June 30 to 217 million pounds (\$403 million) compared with a year earlier. The company's shares rose 7.8 percent this year to 588.5 pence on Oct. 9.

Harcourt Assessment hired a new CEO, Michael Hansen, who took over in July after serving as executive vice president for corporate development at Gütersloh, Germany-based Bertelsmann AG, Europe's largest media company. Hansen, 45, says his company won't slip up again. He blames errors on the enormous demand for made-to-order state tests. "You went from an industry that was largely standard-

How NASD Flunked a Pro

After an eight-year career marketing mutual funds, Linda Cutler says, she felt like a failure when her score came back as a 68 on the Series 7 broker qualifying exam on Feb. 7, 2005. NASD, one of the U.S. brokerage industry's main regulators, requires a person to receive at least a 70 to be a registered representative who can sell stocks, bonds and other securities.

Cutler got the bad news immediately after she took the online test. Two days later, Ryan Beck & Co., a Florham Park, New Jersey-based investment firm, asked her to resign. "I was devastated," says Cutler, 35, who worked out of a midtown Manhattan office. "My employment was contingent on that test."

Because she marketed mutual funds, and not stocks or bonds, NASD, formerly the National Association of Securities Dealers, didn't require Cutler to take the Series 7. Even so, Ryan Beck, which is a unit of Fort Lauderdale, Florida-based BankAtlantic Bancorp Inc., mandates that anyone who dispenses investment advice, including its almost 500 brokers, pass the exam. And Cutler, a senior associate who'd joined the firm in May 2004 after stints at Bank of America Corp. and Morgan Stanley, had spent months studying.

It took Cutler five months to find another

job, this time at New York-based wealth management firm Rochdale Investment Management. She worked as a marketing specialist while she prepared for a retest of the Series 7.

In January 2006, 11 months after being told she'd failed, NASD announced that it had mistakenly flunked 1,882 of the 60,500 Series 7 test takers from October 2004 to December 2005. Cutler was among the wrongly failed.

The error cost Cutler her job, months of salary and her employer's retirement plan contributions. "Plus my reputation," Cutler says. She's suing NASD and the company that scored the exam, Plano, Texas-based Electronic Data Systems Corp., in federal court in Washington. "I lost it all because of that scoring error."

NASD said in a Jan. 6 statement that the 250 questions on each exam are pulled from a large pool of possible items. A scoring program adjusts for difficulties in each question and assigns them a proper weight. The error occurred in the weighting process on as many as 213 questions, the agency said. NASD spokeswoman Nancy Condon declined further comment. Travis Jacobsen, a spokesman for Electronic Data, the world's No. 2 seller of computer services, says it has fixed the problem.

Another test taker, Timothy Wallin, was mistakenly flunked as well. "This was my big

dream," says Wallin of Springfield, Illinois. "When I took the test and failed, it was a crushing blow."

After getting his score, Wallin went to his 2001 Chevrolet Cavalier and spent 15 minutes alone outside the test center. What the hell am I going to do? the 25-year-old asked himself.

Wallin's employer at the time, GCG Financial, a financial services firm based in Bannockburn, Illinois, gave him a second chance to take the test. On Dec. 29, 2004, Wallin passed. Still, he says he was humiliated by the failure. "They should be held accountable," says Wallin, who, like Cutler, is suing NASD and Electronic Data in Washington federal court. NASD and Electronic Data declined to comment on the suits.

Cutler says the grading error is still hurting her. On April 28, she was among the recently hired employees who lost their jobs at Rochdale Investment. She took temporary work before landing a job in September as a project manager for a global financial services firm in New York that she asked not be identified. "If you're in the business to provide test results for people, there should be a level of accuracy," Cutler says.

DAVID GLOVIN and DAVID EVANS

ized to an industry that was highly, highly customized,” Hansen says during an interview in a conference room in his San Antonio office suite, which is adjacent to the test production work floor. “Our most sacred obligation is that the test results are accurate and that they are timely.”

Last year privately held Measurement Inc., a Durham, North Carolina-based test development and scoring company, wrongly failed 890 students out of the 5,461 it tested on Ohio’s high school graduation exams. The company says it scored the exams correctly and then erred when it determined the students’ grades based on the number of questions they answered correctly. “We had a really spotless reputation,” Senior Vice President Mike Bunch says. “This was just devastating to us.”

Pearson Assessments grades 40 million exams each year. The company has the high-profile job of scoring the SAT, which more than 3,000 colleges and universities use as a gauge for admitting students.

Pearson discovered its SAT scoring error in January after two students asked that their results be hand-scored. Score changes affected about 1 percent of the October 2005 test takers, says the New York-based College Board, a nonprofit group that represents 5,000 colleges and oversees the exam. Before most college admission decisions were announced, the College Board re-reported the roughly 4,400 scores that had been underscored. “When you do 12 million tests a year, a lot of people are involved in that,” College Board President Gaston Caperton says. “It’s very hard to get perfection.”

Shane Fulton, a lean youth who played soccer and tennis at George School in Newtown, Pennsylvania, knows the pain of an incorrect score. Fulton had his sights on attending New York University or Lafayette College in Easton, Pennsylvania. In June 2005, at the end of his junior year at the Quaker-run high school, he took his first SAT. He earned a score of 1,910 out of 2,400 on the three-part test, which assesses mathematics, reading and writing. Not satisfied with his performance on the math portion, he took the test a second time in October. He was shocked when the grade came back as a 1,330.

“I knew that something was wrong,” says Fulton, 19, of Yardley, Pennsylvania. He asked to have his exam graded by hand. When the results were returned more than a month later, his score was actually a 1,720, or 390 points higher than initially reported.

By then, Fulton had suffered restless nights, sought sleeping pills from his parents and broken down in tears because of the uncertainty surrounding the scores and his future. Adding to his anxiety, he’d taken the SAT a third time because he didn’t yet know his results on the second test. On that one, he earned an 1,850. “Every year, there’s more of an emphasis on how you do,” says Fulton, who’s attending Northeastern University in Boston and is suing Pearson Assessments and the College Board over the error. “I was thinking I wouldn’t get into any of the colleges I applied to.”

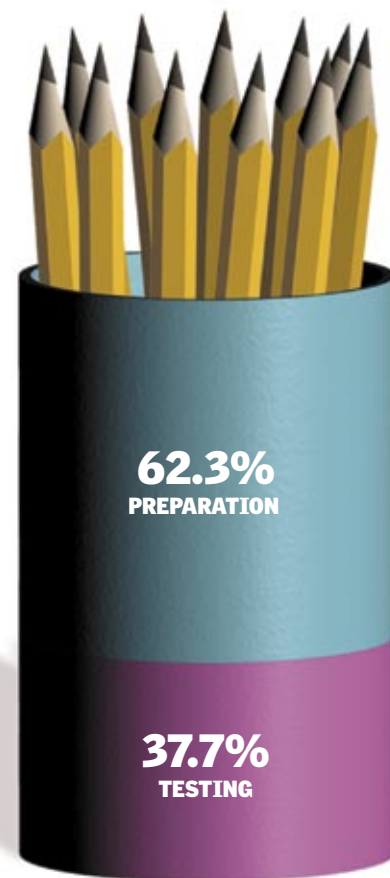
Mistakes may soon cost Pearson Assessments and other test companies business. Educational Testing Service wants to bring scoring in-house to reduce the chance of errors. ETS’s Landgraf

Preparation is everything

Getting students ready for tests is the biggest part of the industry’s revenue.

TESTING REVENUE:

**\$2.8
BILLION**



Source: Eduventures

has directed the company to invest \$50 million so it can expand its scoring operation within three years. He estimates that will produce \$33 million in new annual revenue. Pearson shares gained 11 percent this year as of Oct. 9 to 762 pence.

Having ETS grade his exam didn’t help Pennsylvania teacher Conigliaro, one of the 4,100 false failures on the Praxis test. Forty-four states require the Praxis to evaluate teaching skill and knowledge in a particular field. ETS developed the Praxis and then, in Conigliaro’s case, scored it incorrectly—multiple times.

Conigliaro, 55, an engineer and former machine shop owner, started teaching seven years ago as an intern at Mountain View Junior/Senior High School in Kingsley, Pennsylvania. His employment there was contingent on his passing the Praxis to get final certification. He took the exam in April 2003 and was told he’d failed. He took it again and got a second failing score. He took it a third and a fourth time and again flunked. “I was missing by one or two points each time,” he says.

Conigliaro was fired from his teaching job and wound up working as a bartender. “I didn’t want to leave the house for a year and a half because I was too embarrassed,” he says.

ETS notified Conigliaro in July 2004 that there were scoring errors on his tests and that he had actually passed. In a press release that month the company cited a “statistical anomaly” in the scoring of nine exams from January 2003 to

April '04 and apologized to test takers. ETS spokesman Tom Ewing declined to comment further.

According to court papers by teachers who later sued ETS in federal court in New Orleans, the firm didn't start an investigation of its scoring of short essays until an unnamed state challenged the results. In March, the company agreed to pay \$11.1 million to the test takers to settle the lawsuit.

Conigliaro, who sued and was part of the settlement, says he would have succeeded on at least three of the four exams he was told he'd failed. "Yes, I'm bitter," says Conigliaro, who, after passing the Praxis and getting his license, now teaches business and accounting at Blue Ridge High School in New Milford, Pennsylvania. "I was just about to get tenure, and I had to start all over again."

Test Prep: A+ in Profits, Incomplete in Results

From a storefront office beside a Long Island Rail Road station in Glen Head, New York, about 25 miles east of Manhattan, Rally! Education LLC markets its Test Rehearsal product to two dozen U.S. states. Rally, founded in 2003, is one of as many as 1,800 U.S. test preparation and tutoring firms that are springing up to sell practice materials to help schools get ready for annual No Child Left Behind tests.

Test preparation is booming. Companies took in \$1.7 billion in revenue last year. Profit margins on test prep materials are 20 percent or more compared with margins as low as 3 percent on the year-end No Child exams.

customers who ask that he's both Rally chief executive and an author of the study.

"This is the sort of research that tells us very little," says Richard Allington, a professor of education at the University of Tennessee and past president of the International Reading Association. Further, Allington, who's the author of more than 100 articles and books on reading and education, says constant testing may have a negative effect. "Often, reading gets worse," he says. "The passages they read aren't relevant to the core curriculum, so the kids learn less vocabulary."

Berrent agrees there's no proof that Test Rehearsal alone pushed up scores in New Jersey.

the author of more than 25 books on education, which he sells along with teacher training videos. "Teachers are turning to all sorts of false prophets," he says. "They're being sold a bill of goods."

In Wyoming, Harcourt Assessment is selling a series of prep tests called Learnia. The product, which costs the state \$206,000, has two parts. The first is a group of tests that educators call benchmark assessments. These mid-year exams are designed to tell teachers how much progress a student has made in math, reading, science and writing. The second is a group of what Harcourt calls formative assessments that provide instant feedback.

Cheryl Schroeder, Wyoming's testing director, says Learnia is helping students master the subjects that No Child exams test. Wyoming has started science testing in advance of the 2007-08 federal requirement. "It's how you know the children are making the gains they need to," Schroeder says.

Popham, who serves as a member of the Wyoming advisory committee on testing, says there's no proof that Learnia is helping students. He says the practice tests aren't tailored to Wyoming school standards. "There's no evidence that they're worth a damn," he says.

Harcourt Assessment's new CEO, Michael Hansen, says Learnia exams are under development. "We have not, in any situation, rolled out this product saying 'This is a finished product; here's what it is,'" he says. "They're pilots."

Learnia exams will be customized to test what's being taught in each state, Harcourt spokesman Russell Schweiss says. He says the exams are of high quality.

Popham says one big hang-up in test preparation is that no one is distinguishing the good products, such as those that rely on formative assessment, from the bad ones. "Test publishers are hawking anything they can," he says. "It's absolutely a fraud."

DAVID EVANS and DAVID GLOVIN

'Teachers are turning to all sorts of false prophets,' says James Popham, who has written more than 25 books on education.

"We've been profitable since day one," Rally Chief Executive Officer Howard Berrent says. He says the company has annual revenue of \$5 million-\$10 million.

Rally recommends that schools administer Test Rehearsal practice exams as many as four times a year at a cost of about \$4 per student. Test Rehearsal works, the company says. According to a study that Rally publishes on its Web site, scores in 25 urban New Jersey schools that used Test Rehearsal rose an average of 17.76 percent.

What the study doesn't say is that Berrent, the co-author of the research, is the company's CEO. Nor does it mention that Toms River and other New Jersey districts use prep materials besides Test Rehearsal. Berrent says he tells

"That's the issue with all research," says Berrent, who previously was CEO of Harcourt Interactive Technology, a Harcourt Education unit that developed online classroom tests.

Rick Stiggins, founder of the Portland, Oregon-based Assessment Training Institute, whose assets are now owned by Educational Testing Service, says sophisticated test preparation materials can be helpful. He cites tests called formative assessments, which give immediate feedback on whether students are grasping a new concept. The trouble is, schools get no assurances that the test prep materials they're buying are any good, says James Popham, an emeritus professor of education at the University of California, Los Angeles, and

Errors can occur in the earliest stages of the test-making process and then snowball. In 2003, the Minnesota Department of Education found flaws in questions proposed by Maple Grove, Minnesota-based Data Recognition Corp., a privately held firm that provides testing for eight states. Minnesota school officials reviewed some questions, which are known as items. About 6 percent had no correct answers or multiple correct answers.

“There are other concerns about item quality with another 60–70 percent,” testing director Reginald Allen wrote in 2003 after the company challenged the state’s decision not to renew its contract. The flawed test questions didn’t make it onto state exams. Company lawyer Dwight Rabuse declined to comment except to say that the state later hired a Data Recognition staffer to replace Allen. Minnesota Education Department spokesman Randy Wanke declined to comment. Minnesota now contracts with Pearson Assessments to provide its state tests.

“When you have an education reform agenda that’s relying so heavily on standard tests to ensure school quality, it doesn’t take so many problems to undermine credibility,” says Thomas Toch, co-director of Washington-based research firm Education Sector, who wrote a 2006 report on test errors.

Executives at testing companies say they strive for perfection in the face of state demands for new tests each year, in at least two different subjects and for seven different grades. Stuart Kahl, president and founder of Measured Progress, says the industry uses dozens of quality checks as companies draft, edit, print and deliver exams; retrieve, scan and read papers; and calculate, compare and convert raw scores into

‘I didn’t want to leave the house,’ says a teacher who was fired after his Praxis test was incorrectly scored.

test grades. The process may take two years from start to finish. “There’s no question there are tremendous demands placed on the industry,” Kahl says. “Obviously, when you redo things every year, you have tremendous potential for errors.”

Former Harcourt Assessment President Jeff Galt says state education departments are sometimes to blame for errors that they require their testing contractors to assume responsibility for. He points to Connecticut, which is using Measurement

Spelling optional

CTB/McGraw-Hill hired temporary workers who couldn’t spell to score tests.

The image shows four examples of test forms with handwritten errors circled in red and annotated with red text:

- Form 1:** School Name/Institution: Methodist College. Annotation: School misspelled on application.
- Form 2:** Course of Study (Major/Minor): Phylosophy/Humanity. Annotation: Majors misspelled.
- Form 3:** Major and Minor: in Hungary, pysical education. Annotation: Major misspelled.
- Form 4:** Education Level (Check only one [highest level achieved]): Associate's Degree (checked). Annotation: A bachelor's degree was required to score exams.

Documents digitally re-created by Bloomberg. Source: Florida Education Department

Inc.—its third testing contractor since 2003. The state got rid of Harcourt and then parted ways with CTB/McGraw-Hill. “You have to wonder, Is the problem with the testing company or with the department?” says Galt, 50, who now teaches business at the University of the Incarnate Word in San Antonio. Connecticut Education Department spokesman Henry Garcia declined to comment.

Harcourt Assessment’s inability to follow instructions from Alabama is what cost Jerry Lee Faine Elementary its good name. After the school was notified of its failure to make the required adequate yearly progress, the state placed it in the category of School Improvement, as probation is called under the No Child program. Newspapers publicized the designation, and parents won permission to transfer children to other schools. “People will not move into this community,” says Alfreda Mays-Rogers, whose grandchild is in first grade at the school.

The Alabama Department of Education summoned principal Potter 100 miles north to Montgomery, she says. Officials demanded more teacher training and insisted on additional reading instruction. Potter says she researched curricula used by other schools and dissected years of test data to figure out why her pupils hadn’t passed. Nothing stood out.

During Potter’s crisis of confidence, Kirby Hubbard, the

'When you redo things every year, you have tremendous potential for errors,' a test company executive says.

testing director in Etowah County, about 250 miles to the north, discovered that Harcourt Assessment had miscalculated his schools' No Child results. Harcourt had tallied the scores of students who'd been absent during part of the exam week, failing to follow Alabama's instruction to count the scores of only students who took the entire multipart test, state Education Superintendent Joseph Morton said in a Nov. 8, 2005, letter to Harcourt. That same type of error affected Jerry Lee Faine Elementary. When the state told Potter her school had actually passed on Feb. 9, 2006, she took to the school intercom and made the announcement. Teachers ran into the hallways, cheering. "We were happy, happy, happy," Potter says. "But you turn to the other side, we were mad, mad, mad."

Along with Potter, educators in Florida, Nevada and across the U.S. have to live with test company mistakes every year. Boston College emeritus professor

George Madaus and researcher Kathleen Rhoades say there should be independent oversight of crucial exams. "There's so much error in these products," Rhoades says.

Madaus, co-author of a 2003 study on test errors, envisions an impartial federally financed panel that would monitor state testing programs to ensure they're well crafted and used correctly. Such a board would analyze why there are errors and how they can be minimized. It also may offer a seal of approval on the test preparation products flooding the market, which can generate such a big chunk of a test company's earnings. "This is not anti-testing," Madaus says. "This is an attempt to make testing better."

Potter tries not to be bitter. She notes with pride how her school has now passed the state test for two consecutive years. She has a message for test companies. "They're hurting students more than anything else," she says. "Please don't make that mistake on students. That's a reflection on our school, on my students, on my teachers. That's a reflection on me."

It's also a reflection on the \$2.8 billion test industry, which profits from selling materials to prepare students for high-stakes exams it has a hard time getting right. ▶

DAVID GLOVIN covers the Manhattan federal court at Bloomberg News in New York. DAVID EVANS is a senior writer in Los Angeles.
dglovin@bloomberg.net
daveevans@bloomberg.net

BLOOMBERG TOOLS

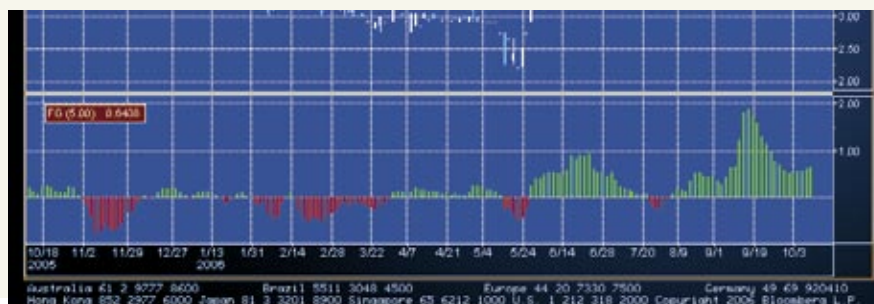
Examining Test Companies

You can use the Related Securities function to find information about educational test units that are subsidiaries of large corporations. For example, type REL LN <Equity> RELS <Go> to find securities and companies related to Reed Elsevier, a holding company with subsidiaries involved in publishing and finance. Click on Major Subsidiaries. Select Harcourt Assessment Inc. for information about Reed Elsevier's test division and some of the brand names it uses.

To apply a proprietary technical analysis study to analyze the stock price of Touchstone Applied Science Associates Inc., a publicly traded test company, type TASA US <Equity> G <Go> 1 <Go>. Select Single Security Historical, enter a name for the study in the GRAPH TITLE field and press <Go>. Define the date

range, and press <Go>. Click on Candle Chart and then on Add Study. Press <Page Fwd> twice, and click on Bloomberg Proprietary Studies. Select Fear Greed (FG) and press <Menu> to display the price graph along with the Fear & Greed indicator, as shown below, which measures the ratio of buying pressure to selling pressure. Green bars show that buying pressure predominates, and red bars indicate that selling pressure is in control of the market.

JOHN DIXON



To graph Reed Elsevier's relative return compared with a benchmark index and its industry group, type REL LN <Equity> RG <Go>.