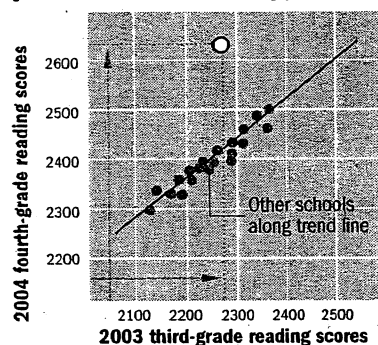


THREE SCHOOLS: OFF THE CHARTS

How to read a "scatterplot" chart

A scatterplot is a chart that shows the relationship between two sets of data. In the charts at right, one set of school scale scores is along the horizontal axis; another set is plotted along the vertical axis. Where the two scores intersect is where the school sits on the chart. As the pattern of dots shows, the two sets of data are closely linked to each other in most schools. The schools suspected of cheating are outliers.

A basic example is shown below, using a hypothetical school that scored 2275 one year and 2620 the following year.

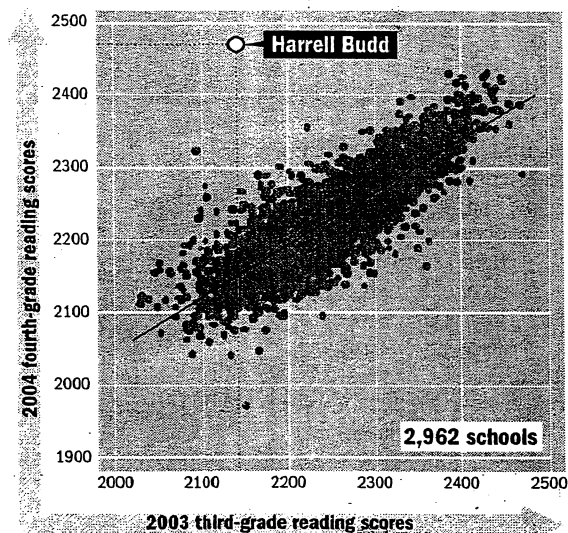


SOURCE: Test scores provided by Texas Education Agency

Harrell Budd Elementary, Dallas

Student stats: 748 students; 94.7 percent poor; 43.3 percent limited English proficiency

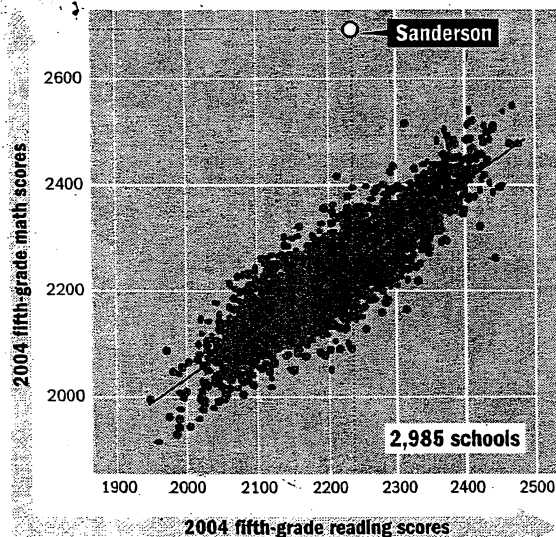
■ Harrell Budd scored poorly in third and fifth grade. But its fourth-grade scores were among the best in the state.



Sanderson Elementary, Houston

Student stats: 365 students; 97.8 percent poor; 14.9 percent limited English proficiency

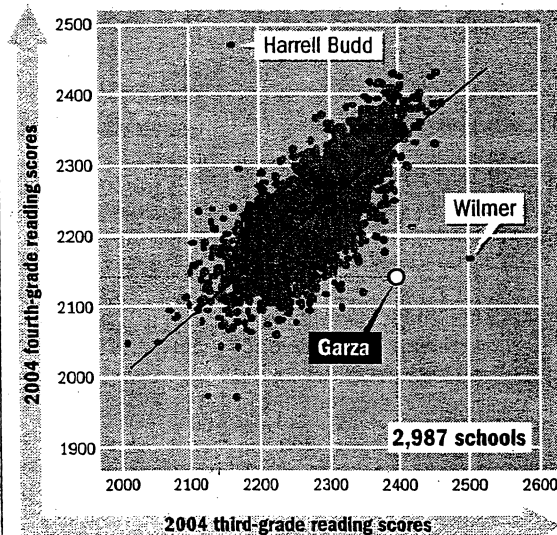
■ Sanderson's fourth-grade math scores were exceedingly low. Its fifth-grade scores were No. 1 in the state.



Garza Elementary, Brownsville

Student stats: 810 students; 99.6 percent poor; 78 percent limited English proficiency

■ Garza's third-grade students, most of whom have problems with English, finished in the top 2 percent of the state in English reading.



Also visible as outliers on the chart: Wilmer Elementary — currently the target of a state cheating investigation — and Harrell Budd

Sunday, December 19, 2004

Cheating may be pervasive

Exclusive: TAKS surges
raise questions about
hundreds of schools

First of two parts

By **JOSHUA BENTON**
and **HOLLY K. HACKER**
Staff Writers

A *Dallas Morning News* data analysis has uncovered strong evidence of organized, educator-led cheating on the TAKS test in dozens of Texas schools — and suspicious scores in hundreds more.

The analysis found a poor urban school where third- and fifth-graders are among the state's weakest readers — but the fourth-graders beat out the state's most elite schools. That's despite the fact that many of its students have trouble speaking English.

It found a desperately impoverished school where the fourth-graders have trouble adding and subtracting — but nearly all the fifth-graders got perfect scores on the math portion of the Texas Assessment of Knowledge and Skills.

And it found schools where in one year's time — if the scores are to be believed — children devolved from top students to barely being able to read.

The News' findings have led to cheating inquiries in three Texas school districts, including the

See **TAKS** Page 14A

TAKS-score

Continued from Page 1A

state's two largest, Dallas and Houston. One of the schools under investigation is a National Blue Ribbon School that a year ago was touted by federal officials as an example of top academic achievement.

"It's very disturbing that this is happening," Dallas schools spokesman Donald Claxton said of data showing unusual swings in test scores at Harrell Budd Elementary in southern Dallas. "There will be a broad-scoped, complete investigation. If there's cheating going on, we want to stop it."

The investigation raises serious questions about the ability of the state's accountability system to reliably measure how schools are performing. The Texas system provided the model for No Child Left Behind, the federal law that measures the quality of all U.S. public schools and punishes those that don't meet standards.

"My sense is that we're seeing a change in culture," said Jim Impara, a former state assessment director in Florida and Oregon. "When you have a system where test scores have real impact on teachers' lives, you're more likely to see teachers willing to cheat."

Houston example

The News' analysis is based on examining scale scores — the little-known numbers behind the passing rates that typically get public attention. The investigation searched for schools with unusual gaps in performance between grades or subjects. Research has shown that schools that are weak in one subject or one grade are typically weak in others.

Take Sanderson Elementary, a school in a poor Houston area.

In 2003, after years of mediocre performance, it reached what has traditionally been the pinnacle for American schools: The U.S. Department of Education named Sanderson a Blue Ribbon School because of rapid improvement in its test scores.

But *The News'* analysis raises questions about the validity of Sanderson's TAKS performance, particularly in fifth-grade math.

Sanderson's fourth-graders scored extremely poorly on the math TAKS test this year. Their average scale score was so low that it ranked Sanderson in the bottom 2 percent of the state: No. 3,173 out of 3,227 schools.

That's roughly what might be expected from a school where almost 98 percent of the student body is poor enough to qualify for free or reduced lunches. Hundreds of research studies have found that student poverty is the single most important factor in student academic achievement.

But Sanderson's fifth-graders had astonishing success on the math test. They had the highest scale scores of any school in Texas, beating every magnet school, every wealthy suburban school and every high-performing school in the state.

Sanderson didn't just finish No. 1. No other school in the state was even close. In scale-score points, the distance between Sanderson and the No. 2 school was as large as the gap between No. 2 and No. 116. More than 90 percent of Sanderson's fifth-graders got perfect or near-perfect scores.

'Educational steroids'

Tom Haladyna, a professor at Arizona State University who studies cheating, said that level of improvement between grades is extremely unusual. He compared it to a weekend duffer beating Tiger Woods by 10 strokes, or a scrub softball player hitting 80 home runs in the major leagues: theoretically conceivable but realistically impossible.

"They're using educational steroids," he said.

Those "steroids" were apparently used only on the TAKS test. Just eight weeks before Sanderson fifth-graders took the TAKS, they took a different standardized test, the Stanford Achievement Test. They didn't fare well, finishing below the national average.

Sanderson's principal, James Metoyer, directed all questions about scores to district officials. Houston Superintendent Abe Saavedra issued a written statement to *The News*.

"At HISD, our credibility and integrity must remain absolutely beyond question," Dr. Saavedra wrote last week. "For that reason, I have asked for a full and thorough investigation of the circumstances surrounding the math scores of this one group of fifth graders."

Dr. Saavedra said the district had reassigned two Sanderson teachers to "other duties" while the district and the state investigate the school's test scores. He also said Mr. Metoyer, the principal, had asked to be reassigned "in order to protect the credibility and the integrity of this investigation."

Dallas school officials reacted similarly when *The News* informed them last week of problems with 2004 test scores at Harrell Budd Elementary.

At Budd, the questions involve the fourth grade, where results in both reading and math were questionable. In the third grade, Budd's students finished in the bottom 4 percent of the state in reading. Not unusual, considering nearly 95 percent of its students are poor and more than 40 percent have limited English skills.

But Budd's fourth-graders were worldbeaters. In reading, they had the second-highest scores in the state, beating schools in Highland Park, Plano and every other high-wealth district. The only school to finish ahead of them was a Houston magnet school for gifted children. Budd's fourth-graders fared almost as well in math, ranking in the top 2 percent of Texas.

After *The News* reported its

findings to district officials, the district launched a cheating investigation at Budd. "We'll find out how extensive the problems are," said Mr. Claxton, the district spokesman. "We're trying to get to the bottom of it."

More than 200 schools

The score swings at Sanderson and Budd were the two most extreme of any of the 7,700 Texas schools whose scores *The News* analyzed. But they weren't the only ones.

More than 200 schools had large, unexplained score gaps between grades or between tests. In statisticians' lingo, these schools had at least one average scale score that was more than three standard deviations away from what would be predicted based on their scores in other grades or on other tests.

In some cases, there may be legitimate explanations for such gaps. School attendance boundaries could have changed dramatically. Or a new public housing development might have radically changed the composition of a school's student body.

But researchers said that large differences between tests are generally signs of something amiss.

"If you see big swings in those numbers, I think we should raise our eyebrows and say this is very, very unusual," Dr. Haladyna said.

The schools most likely to make the list are high-poverty, urban schools, which often feel the strongest pressure to raise scores.

Houston had the most schools

with large gaps: 25 out of the district's 307 schools. Dallas had 21, out of 219 total. Fort Worth had six schools on the list, and no other Texas district had more than three.

Using a stricter standard — four standard deviations from predictions — 41 schools have suspect scores.

The most common pattern involved the third-grade reading TAKS test. Students generally must pass the test to be promoted to fourth grade. That puts more pressure on teachers.

Some examples:

■ Houston's Gallegos Elementary. In 2003, Gallegos' third-graders finished in the bottom 8 percent of the state. In 2004, third-graders zoomed up to the top 2 percent. But the school's reading scores in other grades remained weak.

■ Dallas' Margaret Henderson Elementary, one of Texas' worst schools. It was one of only two North Texas schools to earn the state's "low performing" label from 2001 to 2003. But in 2004, Henderson's third-graders leapt to the state's 73rd percentile in reading. Fourth- and fifth-graders remained in the bottom 5 percent of the state.

Wilmer school

The News began its data analysis in October, when questions were raised about the validity of test scores in the troubled Wilmer-Hutchins school district.

The analysis found strong evidence of cheating at Wilmer Elementary, a long-underachieving

school that rocketed to the best third-grade reading scores in the state. Since the analysis was published, several teachers and students have supported the allegations of TAKS cheating, and the Texas Education Agency has launched an investigation.

In Brownsville, Garza Elementary has scoring patterns similar to Wilmer's. Its fourth- and fifth-graders did poorly on the state's English-language reading test in 2004. Fourth-graders finished in the bottom 11 percent of the state. Fifth-graders were worse: in the bottom 4 percent, 3,336th out of 3,453 schools statewide.

Like Wilmer, Garza teaches the very poor; only three of its 810 students did not qualify for free or reduced-price school lunches. More than three-quarters of its students are considered "limited English proficient" under state definitions.

And, like Wilmer, Garza's students finished in the state's top 2 percent on the third-grade reading test. Almost two-thirds of its students got perfect or near-perfect scores.

Even Brownsville's superintendent thought Garza's third-grade scores were unusual. "I thought, 'That's too good,'" Michael Zolkowski said.

TEA officials are investigating. But district officials have said the inquiry is limited to questions about one or two students' answer sheets, which would not explain the massive score swing.

Researchers differ on how common it is for teachers to cheat. But

most agree it is more common than officials like to acknowledge.

John Fremer, who led the team that developed the new version of the SAT, estimates that between 1 and 2 percent of teachers cheat on their students' behalf on standardized tests. Because those classrooms are spread out among schools, he estimates cheating skews the scores of 3 to 5 percent of schools.

A recent Harvard study of testing in Chicago schools found organized, educator-led cheating in about 4 percent of classrooms, 6 percent when schools with low scores faced consequences.

In an anonymous survey of Arizona teachers by Dr. Haladyna, 11 percent said they improperly helped students on 1991 state tests.

Dr. Impara said that when he started in the testing business in the 1960s, cheating on standardized tests was barely a concern.

"There were almost no stakes attached," said Dr. Impara, who with Dr. Fremer has formed a private test-security company. "The test was intended to provide information on student performance."

Changes in Texas

That started to change in Texas in the early 1990s, with the birth of the state's accountability system. School passing rates were made public and broadcast widely. Schools earned ratings based on their passing rates. The idea: shaming low-performing schools publicly would encourage them to get their ratings up.

Now, in many districts, scores are the key factor in evaluating the performance of superintendents, principals and teachers.

Dr. Haladyna said schools should be able to explain wide gaps in scores if they are not cheating.

"Every time you see one of these schools," he said, "you have the right to ask the question, 'How did you do?' There has to be a program, a method that's producing these results. 'We just tried harder' is not an acceptable answer."

"We just worked real hard" was the explanation given by Geraldine Hobson, principal of Wilmer Elementary, when she was asked last month about Wilmer's astounding third-grade scores. She resigned less than two weeks later.

The News' method of looking for unusual test scores does not catch all cheaters. It does not, for instance, detect schools that cheat consistently across multiple grades and multiple subjects.

It also doesn't catch more subtle cheaters. A teacher who gives students a few correct answers on test day could raise her students' scores enough for them to pass, but not enough for a huge score increase that might draw attention.

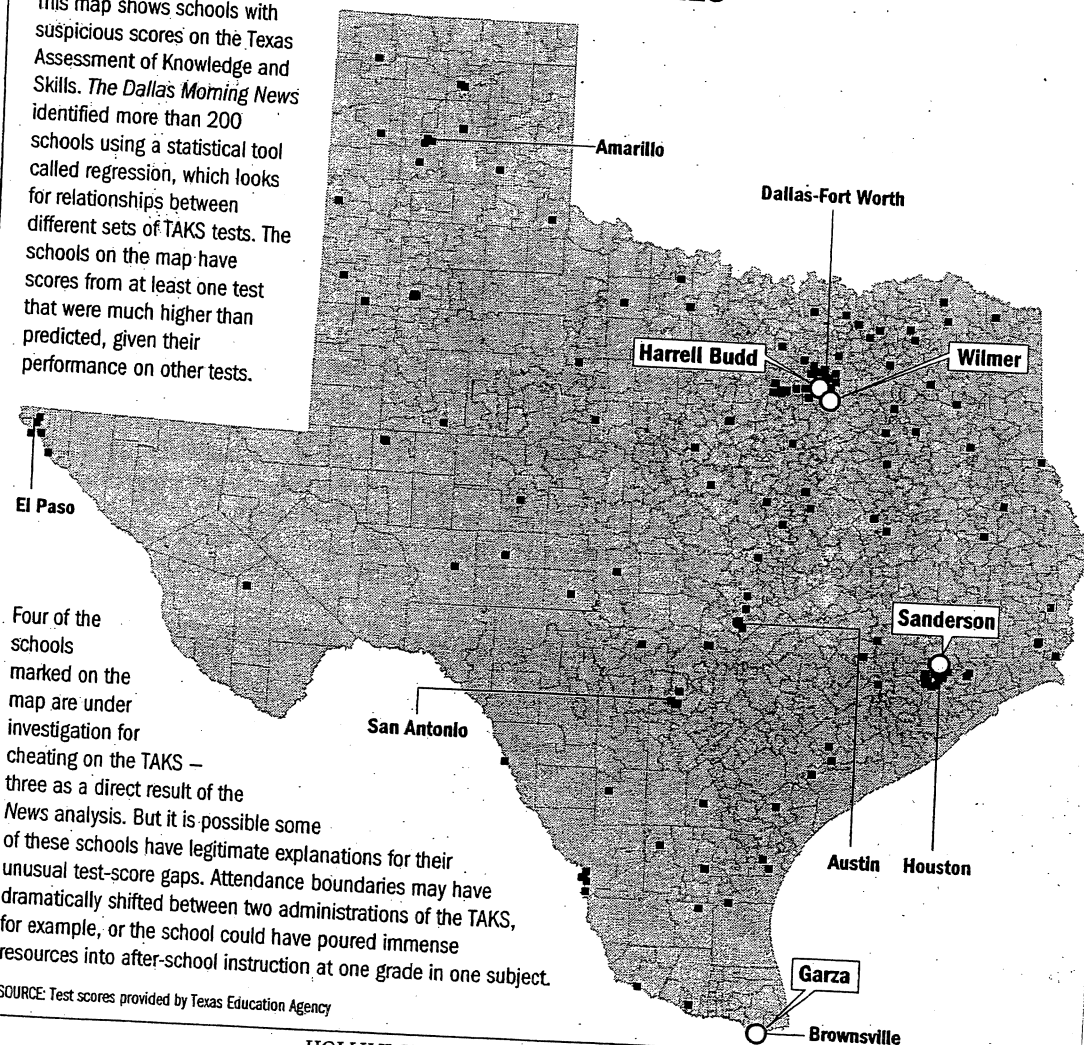
"You're catching the dumb cheaters," Dr. Haladyna said of the analysis. "The smart cheaters you're not going to be able to detect."

Tomorrow: How TEA policies let teachers get away with cheating.

E-mail:jbenton@dallasnews.com
and hhacker@dallasnews.com

TEXAS SCHOOLS WITH SUSPICIOUS SCORES

This map shows schools with suspicious scores on the Texas Assessment of Knowledge and Skills. *The Dallas Morning News* identified more than 200 schools using a statistical tool called regression, which looks for relationships between different sets of TAKS tests. The schools on the map have scores from at least one test that were much higher than predicted, given their performance on other tests.



SOURCE: Test scores provided by Texas Education Agency

HOLLY K. HACKER and JOSHUA BENTON/Staff Writers and CHRIS MORRIS/Staff Artist