

A MIRACLE REVISITED: Measuring Success; Gains in Houston Schools: How Real Are They?

By Diana Jean Schemo and Ford Fessenden

Dec. 3, 2003

As a student at Jefferson Davis High here, Rosa Arevelo seemed the "Texas miracle" in motion. After years of classroom drills, she passed the high school exam required for graduation on her first try. A program of college prep courses earned her the designation "Texas scholar."

At the University of Houston, though, Ms. Arevelo discovered the distance between what Texas public schools called success and what she needed to know. Trained to write five-paragraph "persuasive essays" for the state exam, she was stumped by her first writing assignment. She failed the college entrance exam in math twice, even with a year of remedial algebra. At 19, she gave up and went to trade school.

"I had good grades in high school, so I thought I could do well in college," Ms. Arevelo said. "I thought I was getting a good education. I was shocked."

In recent years, Texas has trumpeted the academic gains of Ms. Arevelo and millions more students largely on the basis of a state test, the Texas Assessment of Academic Skills, or TAAS. As a presidential candidate, Texas's former governor, George W. Bush, contended that Texas's methods of holding schools responsible for student performance had brought huge improvements in passing rates and remarkable strides in eliminating the gap between white and minority children.

The claims catapulted Houston's superintendent, Rod Paige, to Washington as education secretary and made Texas a model for the country. The education law signed by President Bush in January 2002, No Child Left Behind, gives public schools 12 years to match Houston's success and bring virtually all children to academic proficiency.

But an examination of the performance of students in Houston by The New York Times raises serious doubts about the magnitude of those gains. Scores on a national exam that Houston students took alongside the Texas exam from 1999 to 2002 showed much smaller gains and falling scores in high school reading.

Compared with the rest of the country, Houston's gains on the national exam, the Stanford Achievement Test, were modest. The improvements in middle and elementary school were a fraction of those depicted by the Texas test and were similar to those posted on the Stanford test by students in Los Angeles.

Over all, a comparison of the performance of Houston students who took the Stanford exam in 2002 and in 1999 showed most did not advance in relation to their counterparts across the nation. More than half of them either remained in the same place or lost ground in reading and math.

"Is it better or worse than what's going on anywhere else?" said Edward H. Haertel, a professor of education at Stanford University. "On average it looks like it's not." Stanford University has no relationship to the test.

In an interview, Dr. Paige defended Texas's system, saying that it had gradually raised the standards for success over the last 20 years. "Texas measures far more than minimal skills," he said. "The bar is far above what other districts use."

But questions about Houston's accomplishments are increasing. In June, the Texas Education Agency found rampant undercounting of school dropouts. Houston school officials have also been accused of overstating how many high school graduates were college bound and of failing to report violent crimes in schools to state authorities.

The Houston officials strenuously defend the district's record.

Kathryn Sanchez, head of assessment for Houston's schools, said students were doing well on both the Texas exam and the Stanford test, given the city's large number of poor and minority students. Ms. Sanchez said that Houston students had also done well on the National Assessment of Educational Progress, a federally mandated test widely referred to as "the nation's report card."

On that test, fourth graders in Houston and New York outdid children in four other cities in writing, to score at the national average. Fourth graders in New York and Houston also led children in other cities in reading, yet fell short of the national average. Of all six cities, however, Houston excluded the most children with limited English from taking the national assessment, and some researchers suggest that removing such students may have helped raise Houston's score.

But in interviews, Houston school officials acknowledge that the progress in the elementary grades peters out in high school. About 13,600 eighth graders in 1998 dwindled to fewer than 8,000 high school graduates. Though 88 percent of Houston's student body is black and Latino, only a few hundred minority students leave high school "college ready," according to state figures.

Miracle or Mirage?

With its own exam to measure pupil achievement, Texas managed to show educational progress over the last decade on a scale rarely, if ever, achieved before. But as the state's paradigm for school accountability became law for the rest of the nation, the authenticity of Texas's accomplishments has become a major question in education policy.

The Stanford test provides a useful contrast to the state exam, at least for Houston. More than 75,000 students in grades 3 through 8 and grade 10 took the state exam as well as the Stanford test from 1999 to 2002. The Times analyzed performances on these tests, excluding students in special education, and had educational testing experts review the results. The data were obtained under the state's open records act by George Scott, president of the Tax Research Association of Houston and Harris County, a taxpayers group.

"I don't think there was a miracle," said Robert L. Linn, co-director of the Center for Research on Evaluation, Standards and Student Testing at the University of Colorado, who reviewed the calculations. "There were some good positive results, but not extraordinary results like TAAS seemed to show."

The modest improvements in Houston have implications for the national debate. "If you anticipate that you can have the gains shown on TAAS -- and that's what No Child Left Behind would be requiring in many states -- that's not going to be likely to happen, based on this," Dr. Linn said.

The Times analysis of performance on the Stanford Achievement Test and the Texas exam shows this:

*Houston students improved from 1999 to 2002 in most grades, but at only a fraction of the rate portrayed by the state exam. Using a widely employed statistical measure that allows different kinds of tests to be compared called effect size, the gains in the average scores on the Stanford test were about a third of the average gain in the TAAS scores.

*Even students with the poorest skills posted high scores on the Texas test. In reading, a passing score of 70 on the test was the equivalent to scores below the 30th percentile in national ranking on the Stanford test in every grade. In 10th grade, passing the state exam was equivalent to the fifth percentile in the national ranking.

*While the Houston gains on the Stanford test in some grades were large enough to be considered significant in educational testing, the city was not making much headway when compared with national averages. Some 57 percent of Houston students who took the math test in 1999 and 2002, and 51 percent of those who took the reading test, saw their standing relative to children around the country either fall or remain the same.

*On the Stanford tests, the average reading scores for Houston students of all races in grades 9 through 11 have actually dropped since 1999. By contrast, the reading scores for 10th graders on the Texas exam -- the only high school grade in which the state test is given -- showed a large gain over the same period.

*The achievement gap between whites and minorities, which Houston authorities have argued has nearly disappeared on the Texas exam, remains huge on the Stanford test. The ranking of the average white student was 36 points higher than that of the average black student in 1999 and fell slightly, to 34 points, in 2002.

"This says that the progress on TAAS is probably overstated, possibly by quite a margin," said Daniel Koretz of the Harvard School of Education, who also reviewed The Times's analysis, "And when all is said and done, Houston looks average or below average."

Tougher Texas Test

While Texas minority students have made gains on the federal government's mandated national assessment test of reading and math, they were already largely ahead of the average scores of minority students from around the country before the current Texas accountability system began in 1993.

In Houston, the share of college-bound high school graduates that the Texas Higher Education Coordinating Board deemed "college ready" fell to 28.5 percent, or 977 students in 2001, from 33.7 percent, or 1,155 students, in 2000, according to the latest figures available. The board counts only graduates who seek admission to public institutions of higher education in Texas, and says another 10 to 15 percent may seek admission elsewhere.

But many here saw the replacement of the Texas exam last spring with a tougher exam as the most stinging indictment of the test. On the new test, the Texas Assessment of Knowledge and Skills, or TAKS, race gaps widened, and passing rates fell.

Officials here now say that TAAS was only a test of "minimal skills," paving the way for ratcheting up standards with a new exam.

Dr. Paige contends that the TAAS and Stanford tests could not be compared because the Texas test gauges mastery of the Texas curriculum while the Stanford test measures a more general notion of what children should know in a given grade.

But education researchers disagreed.

"These two tests ought to be telling the same story, and they're telling different stories," said Dr. Haertel, of Stanford University.

Dr. Paige also argued that statistical anomalies in the results on the Texas test made comparisons impossible. But testing experts who examined those anomalies said that, if anything, they would reduce the disparities between the two tests.

Watching Children Struggle

In one way or another, Jo Arevelo, Rosa's mother, has watched each of her children struggle through an educational system that was focused tightly on producing high test scores on state exams.

Last summer, Ms. Arevelo tutored her youngest daughter, 10-year-old Angelica, in spelling. Because the state exam does not test spelling, Angelica's teacher never got to it, Ms. Arevelo said one recent afternoon.

Earlier that day, her son, Joseph, took the preparatory exam for the SAT college entrance test, but like many other children that day, he left the exam in frustration -- mystified by vocabulary words like parallelism and euphemism, words he had never encountered in school.

Patricia Anderson, a veteran social studies teacher in Houston, said she was not surprised. Noticing that her high school students could not answer questions after reading passages in their textbooks, she began giving them a vocabulary test at the fourth grade level. Typically, she said, "They flunk it."

"We're all very very frustrated, because all these great scores are coming out of the elementary schools, and when they get to high school it's not happening," Ms. Anderson said. "They do not have the skills they need."

It was not always like this. Many parents welcomed the accountability system that the Houston district pioneered in the 1980's and early 1990's. It was a way, they reasoned, to force schools in poor neighborhoods not to write off their children.

And in some places, it seemed to work, said Rene Barrios, lead organizer for the Metropolitan Organization, a chapter of a group that monitors public services. But in many other places, Ms. Barrios said, the system became the single most important measure of school success and the test itself, for many teachers, became the curriculum. "The whole system has been taken over by the test," she said.

Rosa Arevelo, who graduated from Davis High with a B average, tried to keep pace in college. She made flash cards to help her remember what she studied. She had never learned how to take notes in high school, so at her lectures in college, she took down everything the teacher said.

Her textbook looks as if it is filled with neon lights: entire paragraphs are highlighted in bars of bright pink and yellow. In the unrelenting array of information, she could not tell what mattered.

"When you get to college," she said, "you're just supposed to know. But nobody ever taught us."

How the Houston Test Scores Were Analyzed

The calculations for this article were based on the records of 75,000 Houston students in Grades 3 through 8 and Grade 10 who took the Texas Assessment of Academic Skills as well as the Stanford Achievement Test in 1999 and 2002. The New York Times calculated the change in the mean scores in math and reading for those two years for each grade and divided by the standard deviation for 1999, a recognized method for calculating the effectiveness of new teaching methods known as effect size. The method allows different kinds of tests to be compared.

The national rankings that were equivalent to the passing score of 70 on the Texas exam were calculated with a regression equation, a statistical measure that uses all student scores to predict the result on one test from the scores on the other.