

**NEWS**

# Are Texas Kids Failing? Or Are the Tests Rigged?

Leading educators and testing experts have studied the exams and find that they are asking kids to understand passages aimed two or more years above their grade level.

---

**BY** MIMI SWARTZ

---

**ISSUE** APRIL 2019

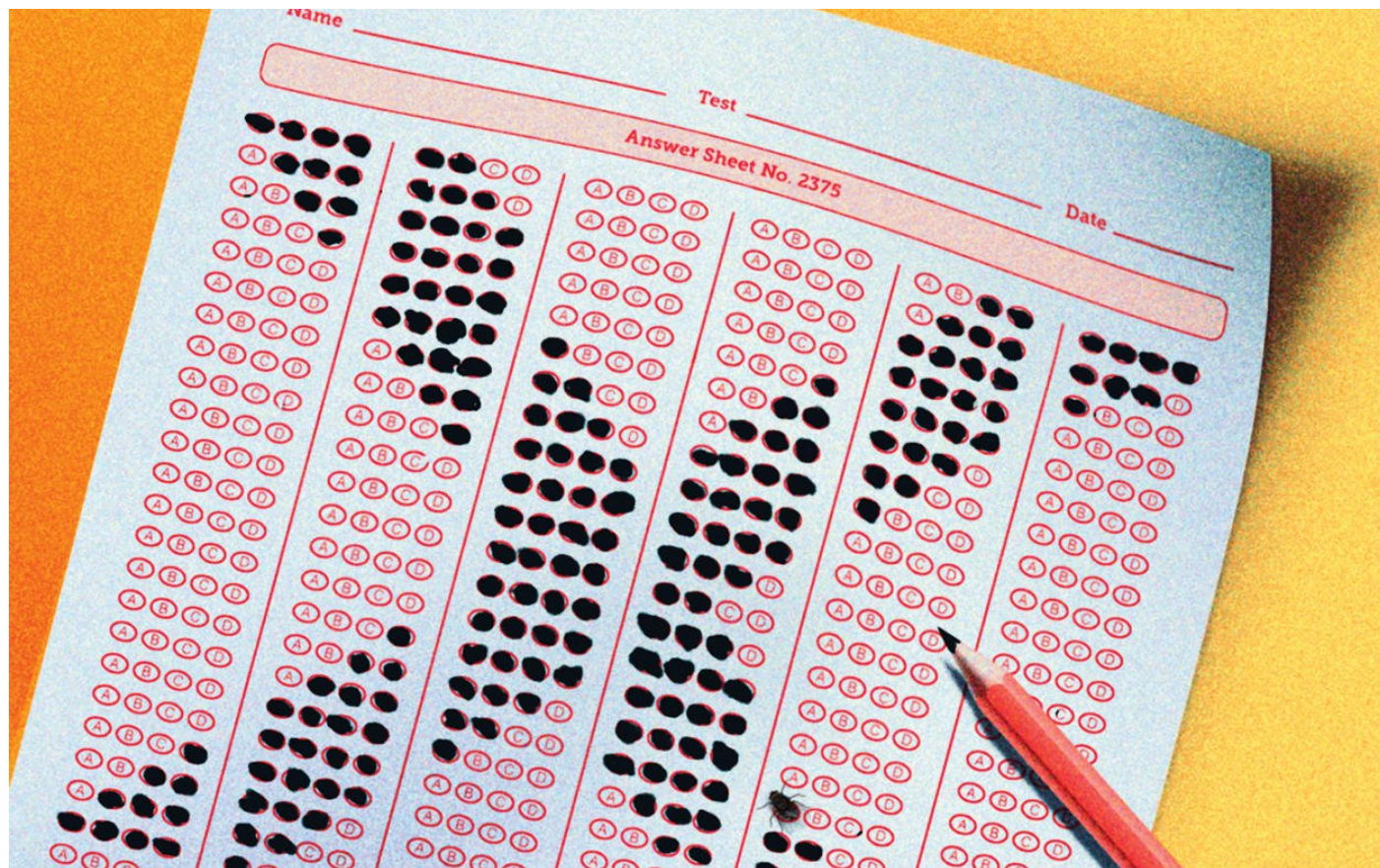
---

**SHARE**

---

**NOTES** **50 COMMENTS**

---



*Illustration by Mark Pernice*

Cynthia Ruiz is the kind of English teacher everyone wishes they had. A youthful 39, she has an embracing manner and a contagious enthusiasm for learning. She began teaching right after college graduation, in the spring of 2003, and in 2012 she took a job at Connally High School, in Pflugerville, just north of Austin. A world away from the capital's hipsters, software moguls, and lobbyists, Connally's students were, in many cases, poor and had learned English as a second language. It didn't surprise Ruiz that most of them hated to read.

Ruiz was determined to change their minds. She had grown up in circumstances similar to those of many of her students; after her father lost his job, she had to take out loans to put herself through college and graduate school. She knew that if she didn't get her kids reading, they'd never have a shot at better lives.

Ruiz had her students read required books like *Of Mice and Men*, but she also let them have free reading time, when they could open up any book in the classroom library she had spent more than \$400 out of her own pocket to build. When students complained that they couldn't find anything to read, Ruiz asked about their lives. One kid mentioned that he watched the cable series *Shameless* because, like the characters on the show, he too had a tough family life. Ruiz gave him a copy of *The Glass Castle*, Jeannette Walls's memoir of her itinerant life with dysfunctional parents. "It was like reading his own life," Ruiz recalled. "Over the next few months, his attitude and performance changed completely."

But the same year Ruiz started at Connally, the state of Texas began its switch from a standardized test known as the TAKS, the Texas Assessment of Knowledge and Skills, to one that was supposed to be more rigorous, the State of Texas Assessments of Academic Readiness, or STAAR. Freshmen that first year had to take five tests over twenty hours on five days to demonstrate their readiness for sophomore year: English I reading, English I writing, world geography, biology, and algebra. The STAAR was what was known as a "high stakes" test—results could determine whether a student could advance to the next grade and, ultimately, graduate.

There were problems from the start. Teachers received little preparation for the new exam, and once the kids started testing, their reading scores in particular were surprisingly low. Ruiz had language arts classes full of kids reading at grade level who couldn't pass the STAAR test. She tried to get information from the Texas Education Agency that might explain the low scores, but she got no response. She felt like she had no way to help her kids do better.

Over the next few years, Ruiz watched as her students repeatedly failed the STAAR test. “I saw a huge disconnect and disengagement,” she said. Statewide, 71 percent of all students who took the English I test in the fall of 2016 failed.

With the low scores came demands from the district for more test preparation in class. More and more students were now spending their time in cram sessions and remedial classes because passing the STAAR was a significant requirement for graduation—and low STAAR scores reflected badly on the school. “I had kids drop out because they didn’t see light at the end of the tunnel,” Ruiz said. Some students felt it was safer to go for a GED or an online diploma than to risk failing the STAAR test and finding themselves unable to graduate.

By 2018, burned out and exhausted from “drill and kill” teaching, Ruiz quit. “Not one college or employer looks at these scores, and we are spending millions of dollars on them,” she says. “And for what?”

Ruiz is not alone in her frustration. Over the past seven years, something strange has been happening in Texas classrooms. Accomplished teachers whose kids were reading at grade level by virtually all other measures were seeing those same students fail the STAAR test. School nurses began to see kids suffering from anxiety just before test days—kids who complained of stomachaches, headaches, and just plain old fear—in numbers far greater than they had during the TAKS era. Parents were desperate to find out why their once high-performing kids were suddenly stumbling and feeling despondent about their education. And student motivation wasn’t helped by curricula increasingly dominated by stultifying test prep.

The trouble began in 2012, when the STAAR test replaced the TAKS, which Texas officials considered too easy. The Texas Association of Business, worried that Texas high school students weren’t prepared to enter the workforce, was particularly influential in pressing for a more challenging exam.

The STAAR test is used not just to evaluate student progress; the scores have also been used to evaluate teachers, individual schools and principals, school districts, and, by extension, the entire enterprise of public education in Texas. And many politicians in Texas frequently cite underperforming schools as evidence that the state should offer vouchers that would enable parents to send their kids to private schools. Governor Greg Abbott, among others, has noted that according to the STAAR test, only 40 percent of Texas third graders are reading at grade level.



The argument in favor of tough testing is a simple one and one that many people on both sides of the ideological divide agree on: Texas has to get its kids and its public schools up to the highest standards if we want to have the educated workers and informed citizens we need in the twenty-first century. The problem is, no one knows exactly how to make that happen, and everyone has an opinion.



**Students at Dyess Elementary School, in Abilene, on their way to take the STAAR test on March 24, 2016.**

*Nellie Doneva/The Abilene Reporter-News/AP*

**There were grumblings** about the STAAR test early on. Months before it was actually administered the first time, Susan Szabo and Becky Sinclair, professors of curriculum and instruction at Texas A&M University–Commerce, published a report in the academic journal *Schooling* titled “STAAR Reading Passages: The Readability Is Too High.” Their research, based on sample STAAR questions that were made available before the test’s debut, suggested that the STAAR test didn’t accurately measure whether students were reading at grade level. Their examination of five different readability tests—commonly used academic measures that rate the appropriateness of written passages for various grade

levels—showed that in order to comprehend various STAAR reading test passages, most students would have to be reading at higher than their grade level. A third grader, for instance, would have to comprehend at a fifth-grade level.

Szabo and Sinclair's paper made no waves. The STAAR test was new, and no one in power heeded the warning implicit in their research. But some parents were alarmed by what they were seeing during the first years of the STAAR testing regime; during the 2013 legislative session, there were protests against the exams at the Capitol. At the end of the session, the Legislature eased some of the most onerous requirements; the number of exams a high school student would have to take was reduced from fifteen to five—English I, English II, algebra, biology, and U.S. history.

Then came another study, in 2016, by Michael Lopez and Jodi Pilgrim, a graduate student and a professor of education at the University of Mary Hardin-Baylor, in Belton. They used six different readability tests to evaluate the STAAR reading test—the five Szabo and Sinclair had used and the Lexile scale, which is regarded nationally as the standard gauge of any publication's degree of difficulty (libraries use the Lexile scale to direct kids to age-appropriate books). Like Szabo and Sinclair, they determined that the STAAR test contained passages that were too difficult for the targeted age groups, confirming what many teachers were seeing in their classrooms.

#### **IT'S NOT JUST TEXAS**

**Researchers wrote, of New York's exams, "We conclude that testing instruments that put children in a virtual stupor cannot be defended as sound testing practice."**

At the same time, problems with the administration of the test became apparent. The Princeton, New Jersey–based Educational Testing Service had been given a four-year, \$280 million contract, but that enormous sum didn't keep it from misdelivering tests to schools, losing records of test answers, and foot-dragging on reporting scores. So it was no wonder that, three years ago, a group of fifty Texas school superintendents took their complaints to the Texas Education Agency, which oversees the STAAR test. The TEA did little to nothing in response.

In 2017, however, some headway was made. One group, Texans Advocating for Meaningful Student Assessment, convinced the House to eliminate some of the harsher aspects of the test, including the use of the STAAR as a determinant of whether a child in eighth grade or lower advanced to the next grade. "We've heard the stories of third graders throwing up the

day before the test because they were just physically ill. This eliminates that pressure on those students,” the author of the bill, Representative Gary VanDeaver, told KHOU News, in Houston. (The bill later died in the Senate.)

It’s easy, especially in Texas, to explain away some of the complaints as just so much whining from educators who don’t want to admit that the schools they run aren’t up to snuff. After all, according to data compiled by the journal *Education Week*, our state ranks fortieth in education quality. But many people—including many of those who are most concerned about student outcomes—believe that the STAAR test is simply too flawed for its results to be used as Exhibit A in the case against Texas schools.

**H. D. Chambers is one** of a growing list of educators who think the STAAR test has done tremendous damage to the Texas educational system. He’s the superintendent of the Alief Independent School District, southwest of Houston, and also the president of the Texas School Alliance, an organization that represents many of the largest school districts in the state. A circumspect man with pale-blue eyes and a dry wit, he leads one of the poorest school districts in Texas. He knows from low scores.

But before 2012, TAKS reading scores at Alief were slowly rising. Since the inception of the STAAR test, though, he has seen scores flatline, no matter how much additional exam prep his kids are subjected to. Chambers is skeptical of these numbers. He knows that his teachers and students are working harder and smarter to get the scores up. “Based on the many reading and literacy experts who have spent years addressing the issue of literacy, far more children are reading at or above grade level than the number the state is publishing,” he said. “No one, including me, is saying it’s a hundred percent, but it’s a lot higher than the forty percent some claim.

“I want to be clear and emphasize that this issue is not an attempt to lower standards or expectations,” Chambers said. “We are trying to align the standards and what teachers are told to teach with what is tested and how those results are applied to accountability. Every reading and literacy expert who has studied our concerns can’t be wrong on this. This is not anti-testing. This is not anti-accountability. We just want the truth.”

One of his concerns, voiced by many others, is the Texas Education Agency’s lack of transparency, its failure to forthrightly acknowledge that it has radically raised the bar for schoolchildren. “If the decision was made to test kids in reading passages that are above their grade level, everyone needs to know that,” Chambers said. One parent who has campaigned against the STAAR explained the situation using—inevitably, given that this is



Texas—a football analogy: “In football, you get to the end zone, and you score a touchdown. So what happens if the referees get together and decide you have to get *past* the end zone, but they don’t tell the players or the coaches that? That’s kind of what TEA has done.”

The Texas School Alliance and several state and national testing experts met on February 11 with Texas Commissioner of Education Mike Morath at the TEA headquarters, three blocks north of the Capitol. Wielding the latest findings on the misrepresentation of student achievement by the STAAR reading test, they argued that the test was out of sync with numerous readability tests and the results were hurting schools, teachers, parents, and kids.



**Commissioner of Education Mike Morath speaks at the Capitol, in Austin, on October 12, 2017.**

*Jay Janner/Austin American-Statesman/AP*

Morath, an Abbott appointee who took over as head of the TEA in 2016, is a technology entrepreneur and firm believer in data who fought for excellence in the school board trenches of the Dallas Independent School District. He is considered to be smart, and sensitive to the plight of underprivileged students, but also stubborn when convinced he’s

correct. Most educators give him high marks. “I think he cares about the kids and he’s trying to do what’s right,” said Chambers.

But Chambers and his group didn’t feel that Morath gave them the hearing they had hoped for. Though the meeting (including a subsequent one-on-one between Chambers and Morath) stretched to more than three hours, they claim that Morath responded to their concerns with a lot of jargon and refused their bottom-line request: to reevaluate the way the STAAR reading test is being administered. Morath told them that the state had its own indicators that showed that the results were correct, but he declined to share that information during the meeting. The agency had looked into this issue before, he said. He wasn’t going to do it again.

The next day, though, the TEA shared with the TSA and other groups a redacted version of the study that it claims vindicates the current testing regime. Yet Thomas Ratliff, a former member of the State Board of Education who now lobbies for the Texas Association of School Boards, said that the study essentially proves much of what the test’s critics have said; it acknowledged, for example, that the hardest third-grade questions could be considered more appropriate for students who are reading at a fifth-grade level. (Though some of the easiest third-grade questions were rated as appropriate for below grade level.)

Morath did not respond to our request for an interview, but Jeff Cottrill, the TEA’s deputy commissioner of standards and engagement, explained that the agency’s research on the STAAR reading test included early reviews by Texas teachers and students. “The test is rooted in Texas standards and reviewed by Texas teachers and field-tested by Texas students,” Cottrill said. “I have to tell you, the process by which TEA determines what goes in this test is solid.”

Though critics dismiss that method as nothing more than a gut check, Cottrill defends the process’s integrity. “TEA relies much more on people to assess the quality of the test than computer-based algorithms,” he explained. “Some Dr. Seuss books are actually written at a higher Lexile than *The Grapes of Wrath*.” (That’s technically true: the Dr. Seuss books designated for adults to read to children, such as *The Butter Battle Book*, may in fact present challenges to young readers. But *The Cat in the Hat* and *One Fish, Two Fish, Red Fish, Blue Fish* do not score higher on the Lexile scale than John Steinbeck.) Chambers says that new research conducted at A&M will be released in the next few months and will show that, according to the latest STAAR test, even fewer kids are reading at grade level today than in 2012—a result, he believes, that conflicts with virtually everything he knows about what has happened in Texas schools over the past few years.



As is usually the case with education conflicts, while adults argue, it's the children who suffer most. Ratliff said that, according to the numbers he's seen, the reading levels of 25 to 30 percent of Texas schoolkids—1.25 million or so—are misidentified. And that sort of failure has a domino effect. "Think about its effect on the economic engine of Texas," he said. "The concentric circles of damage range from mental and psychological damage to schoolchildren to falling real estate values to our ability to recruit businesses. I've tried to get my arms around the damage, and I can't."

On March 5, at the request of Chairman Dan Huberty, the House Public Education Committee held hearings on the STAAR test that lasted more than six highly emotional hours. The main topics were the tests' readability level and the pressure put on kids because of the tests' high stakes. Morath was there with three backup experts defending the exam, but the opposition was out in force. Everyone from testing experts to superintendents to weeping parents gave accounts of kids crying, vomiting, and locking themselves in the bathroom at school to avoid the STAAR. They begged the legislators to make changes in the test.

So perhaps the STAAR's critics will get some satisfaction during this legislative session. "If we're so focused on accountability," said Cynthia Ruiz, the former Pflugerville high school teacher, "I'd like the pressure taken off students and teachers and more accountability placed on the TEA."

*This article has been updated since it was originally published online. The version you are reading here appears in the April 2019 issue of Texas Monthly with the headline "STAAR Wars." Subscribe today.*

---

***If you appreciated this reporting and want to support our future work on education, politics, and more, subscribe today for a \$1/month digital subscription or \$1.25/month digital + print subscription.***

---

3/25/2019

Are Texas Kids Failing? Or Are the STAAR Tests Rigged?

**50**

COMMENTS

TAGS: EDUCATION, LEXILE SCALE, MIKE MORATH, SCHOOL TESTING, STAAR