

<https://www.texasmonthly.com/news/texas-kids-failing-staar-tests-rigged/>

from Texas Monthly

Are Texas Kids Failing? Or Are the Tests Rigged?

A group of 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

DATE

FEB 20, 2019

Over the last few years, something strange has been happening in Texas classrooms. Accomplished teachers who knew their kids were reading on grade level by virtually all other measures were seeing those same kids fail the STAAR, the infamous State of Texas Assessments Readiness test.

The effect on students was predictable: kids who were diligently doing their homework and making good grades in class were suddenly told they were failing in the eyes of the state, which wasn't so great for their motivation. Parents were desperate to find out why their once high-performing kids were suddenly seen as stumbling. Teachers felt like failures, too, but had no idea what they were doing wrong, after years of striving to adopt practices proven in successful schools across the country. What's more, the test results were quickly weaponized by critics of Texas public schools, many of whom advocate state-funded vouchers that would allow parents to send their kids to religious and other private schools.

The stakes of such exams are perilously high. The STAAR test, developed by the Educational Testing Service in Princeton, N.J., had replaced one provided by the British firm Pearson, which Texas officials considered too easy. The STAAR test is used to evaluate students, teachers, individual schools and principals, school districts, and, by extension, the entire enterprise of public education in Texas. Fifth and eighth graders who fail the test can be forced to repeat a grade; high school students may not graduate if they don't pass three of the five STAAR year-end exams.

On its face, this approach makes sense. This is, after all, the Age of Accountability, and, according to Governor Greg Abbott and other prominent state leaders, only 40 percent of Texas third graders are reading at grade level. The STAAR numbers are cited as positive proof of that. 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. There isn't a minute to lose.

This reasoning may explain why a report issued in 2012 by two associate professors at Texas A&M was overlooked. Called "STAAR Reading Passages: The Readability is Too High," by Susan Szabo and Becky Sinclair, the report suggested that questions on the STAAR test were too hard to accurately measure whether students were reading at their grade level.

The researchers' examination of five different "readability tests"—commonly used academic measures that rate the appropriateness of written passages for various grade levels—showed, for instance, that in order to comprehend various passages, a third grader would have to read on a fifth-grade level. A fifth grader would have to read on a seventh-grade level, and so on. Generally, the testing showed a gap of about two years. Szabo and Sinclair's paper made no waves. The STAAR test was new, and if there was a warning included in the research, no one in power thought to consider it. An organization called Texans Advocating for Meaningful Student Assessment lodged protests, but they were rebuffed.

Years passed. The STAAR reading test reported more failures and stirred more concerns. Teachers and administrators continued to see that the STAAR scores didn't "align" with other indicators of reading levels. Specifically, the numbers didn't match those of 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.)

In 2016, another study was released, this time by Michael Lopez and Jodi Pilgrim, two professors at the University of Mary Hardin-Baylor, in Belton, Texas. They, too, found that readability formulas showed that the STAAR test contained too many difficult passages for the targeted age groups—"materials may be problematic for teaching and learning"—which confirmed what many teachers were seeing in their classrooms. That same year, a group of fifty Texas school superintendents lodged their protests with the Texas Education Agency (TEA), which administers the STAAR test.

It's easy, especially in Texas, to explain away some of the complaints as just so much whining. According to recent Education Week studies, our state ranks 40th in education quality. The blame for our sad showing has been placed on allegedly unqualified and unaccountable teachers, uninvolved parents, and corrupt administrators and school boards.

But what if that showing isn't as sad as we've been told? What if the STAAR test isn't measuring what it says it's measuring: i.e., that a third grader is reading at a third-grade level, rather than a fifth-grade level?

H.D. Chambers thinks that's exactly what's happening. He's the superintendent of the Alief school district southwest of Houston and also the president of the Texas School Alliance (TSA), an organization that represents the largest school districts in the state. A circumspect man with pale blue eyes and a very dry wit, he leads one of the poorest and most diverse school districts in Texas. He knows from low scores. But before the STAAR test, he was seeing that reading scores at Alief were slowly rising. Afterward? Scores flatlined. He was skeptical. Chambers knew that his teachers and students were working harder and smarter. "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 100 percent, but it's a lot higher than the 40 percent some claim."

The TSA, along with testing experts with whom it consulted, pushed for a meeting with Texas Commissioner of Education Mike Morath, to show him their latest findings on the misrepresentation of student achievement by the STAAR reading test. "I want to be clear and emphasize that this issue is not an attempt to lower standards or expectations. 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," Chambers said. The STAAR test is supposed to measure what kids learn during a given school year, not their overall knowledge. Schools need to be accountable, certainly, said Chambers, but "it's vital that we have test questions that accurately measure what they say they are measuring at any given grade level."

Morath agreed to a meeting on February 11 with Chambers and several state and national testing experts, including Dee Carney of Moak, Casey & Associates, an Austin firm that consults on school finance and assessment issues. Chambers and his group urged Morath and his deputies to take a fresh look at the accuracy of the STAAR test. The group met in a conference room at the monolithic TEA

headquarters, three blocks north of the Capitol. The TSA group argued that the reading test was “misaligned”—that it was not grade-level appropriate. It was out of sync with the basic Lexile findings in particular, they said, and the results were hurting schools, teachers, parents, and kids.

Morath, who is 41, is a slight, balding technology entrepreneur who proved himself in the school board trenches of the Dallas Independent School District. He was appointed commissioner of education by Governor Abbott in 2015. Morath is considered to be smart, sensitive to the plight of underprivileged students, and also quite stubborn. “I think he cares about the kids and he’s trying to do what’s right,” said Chambers. Morath did not, however, give the group the open-minded hearing they had hoped for—or at least that’s how his audience perceived it. They felt politely dissed, complaining that Morath responded with a lot of jargon and refused to reevaluate the way the reading test is being administered. He claimed that the state had its own indicators that showed the results were correct, but he declined to share that information. The agency had looked into this issue before, Morath said. He wasn’t going to do it again. “They have a study that they claim justifies” how the STAAR test is being administered, said Thomas Ratliff, a former member of the state board of education. “We have requested that study and we have not seen it.”

Morath did not respond to our request for an interview, but we were able to speak with Jeff Cottrill, TEA’s deputy commissioner of standards and engagement. He explained that TEA’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.” Critics dismiss that method as nothing more than “a gut check,” as none of the test passages were run through standard readability measurements such as the Lexile. Cottrill confirmed that the test was not sent through a Lexile analysis. “TEA relies much more on people to assess the quality of the test than computer based algorithms... Some Dr. Seuss books are actually written at a higher Lexile than *The Grapes of Wrath*,” he said.

The Lexile scale was not the only readability test by which researchers outside the TEA have evaluated the STAAR reading test. Dee Carney, the Austin testing expert, pointed out that the A&M research used five readability studies and the Mary Hardin-Baylor research used six. Chambers says new research conducted at A&M is to be released in the next few months and shows even more misalignment, or failing kids, today than in 2012. “If the decision was made to test kids in reading passages that are above their grade level, everyone needs to know that,” Chambers said. “If a

third grade reading test is meant to determine if a student is reading at the third grade level, then the test questions should be based solely on what was taught in [and before] third grade, not what might be taught in the fourth, fifth, sixth, or seventh grade.”

The consequences, Chambers said, can be severe. “To me, here is the bottom line: if Texas expects every third grader to read like a fifth grader or every fourth grader to read like a sixth grader, then we all need to be prepared to see lower performance. Based on all the expert information that has been provided, these unrealistic standards have the potential to destroy learning.”

One step toward addressing issues with the STAAR test would be to hold public hearings. Two Republican state legislators from the Houston area say they want to do so, says Ratliff, but so far none are scheduled. One of those lawmakers is Larry Taylor, who chairs the Senate committee on education. The other is Dan Huberty, who chairs the House committee on public education.

As usual with education conflicts, while adult officials argue, it’s the schoolchildren who suffer most. But they’re not the only ones. Ratliff, the former state board of education member, estimates that 25 to 30 percent of Texas school kids are misidentified as reading below grade level—1.25 million or so children. “Think about its effect on the economic engine of Texas,” he said. “The concentric circles of damage ranges 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.”

Added Chambers: “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.”

This article has been updated to correct Commissioner Morath’s age (he is 41, not 43), and to correct a quote from TEA deputy commissioner Jeff Cottrill.