

The Washington Post

[Answer Sheet](#) Analysis

# Are these standardized test questions too hard for kids?

---

By [Valerie Strauss](#)

Texas Monthly published a story — “[Are Texas Kids Failing? Or Are the Tests Rigged?](#)” — which raises questions about whether questions on the state’s high-stakes standardized English Language Arts tests are written above grade level for many children. One Texas lawmaker, state Rep. Mary Gonzalez (D), quickly called for an investigation into the exams.

The article cites a [2012 report](#) by two associate professors at Texas A&M University at Commerce, who analyzed English Language Arts exam reading passages and found that many were written at least two grades above grade level.

This is not a story unique to Texas, nor are the consequences of poorly worded, culturally biased or otherwise inappropriate questions on high-stakes tests. We’ve heard, for example, about problem questions for years from different states, including the infamous “talking pineapple” questions on a 2012 New York state test. Problems in New York didn’t stop, as you will learn below.

The Texas Monthly story starts this way:

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 of Academic 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, New Jersey, 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.

Let's move to New York.

From 2012 to 2016, the state had a \$38 million contract with the London-based education company Pearson for tests aligned with the Common Core State Standards to be given to some each year. New York did not renew the contract after repeated problems. Questar Assessment won its own five-year contract, for \$44 million, and has completed two testing cycles. It will implement its third in the spring.

Testing companies are not usually required by states to provide extensive data to the public about their exam programs and are not required to seek or even allow independent review of its questions. With that in mind, testing expert Fred Smith and Robin Jacobowitz of the State University of New York at New Paltz, have analyzed reading passages along with the accompanying constructed response questions. They looked at questions from Pearson-developed tests that were made public by state officials on their EngageNY website.

Smith submitted Freedom of Information Act requests to the state for test scores and was then able, in some cases, to match the released passages against scores students received in response to the questions. This data is at the heart of the analysis below, which focuses on the harsh impact two specific passages had on all children in the state and especially New York City's English Language Learners, special education students, and black and Hispanic children (who make up 68 percent of the city's test population).

Smith is an expert on standardized testing and a retired administrative analyst with the New York City public school system. Jacobowitz is director of education projects at the Benjamin Center at SUNY at New Paltz, where she taught for years in the School of Education.

It is true that in both of the studies mentioned above, in Texas and New York, test questions examined were from several years ago. There is no reason to think, however, that much has changed since then.

### **By Fred Smith and Robin Jacobowitz**

In 2018, we published a report titled "[Are Turning Our Kids Into Zeroes: A Focus on Failing](#)," which strongly suggests that many students have been unable to understand readings and write intelligible answers to Common Core-based questions about them on the statewide English Language Arts (ELA) tests.

Parents and teachers have complained to no avail that reading passages on the tests were developmentally inappropriate, particularly for third- and fourth-graders. Now we suggest that the critics seem to have been right.

We recently looked at the overall impact of the statewide exams on the 1.2 million students who take the tests each year. This included separate analyses for the 440,000 children in New York City, or 37 percent of the state's test population. These tests were prepared by Pearson Inc., the company that used to have a contract with the state but no longer does.

We showed that these exams had the most dire effect on 8- to 9-year-old children right after the state introduced tests aligned with the Common Core State Standards in 2013. For third graders, the switch to Common Core-aligned exams resulted in a surge from 11 percent of students who got zeroes — meaning their answers were deemed entirely incomprehensible — to 21 percent. And for fourth graders, the jump was from 5 percent to 15 percent. (See score chart below.)

English Language Learners, students with disabilities, and black and Hispanic students were particularly hard hit. This was clear from the data we obtained from the New York City Department of Education, which allowed us to analyze the tests' impact on each of these groups.

We wrote follow-up blog posts about the [broad impact the tests had](#) and the specific [impact they've had on minority and English-as-a-second language](#) learners.

Now let's look at the readings and test questions that stumped our kids. First read the following two passages. The questions are after each passage. Remember these passages and questions were given to 8- and 9-year-old students.

---

**GRADE 3, 2014, SCIENCE FRICTION, QUESTION 45**

***Question: Why is the setting of the story important? Use two details from the story to support your response.***

**-0-0-0-0-**

**Grade 4, 2015, Hattie Big Sky, question 45**

***Question: How are the chickens presented as characters in "Excerpt from Hattie Big Sky"? Use two details from the story to support your response.***

These are questions from Common Core-aligned, Pearson-designed tests that have the furthest outlying zero scores from students.

"Science Friction" appeared on the 2014 Grade 3 test; this passage and question No. 45 formed the most incomprehensible combination students faced that year, with 48 percent of New York State students scoring zeroes on it. A year later, in Grade 4, "Hattie Big Sky" was nearly as much of a stumper, with 41 percent zeroes.

But these aren't isolated examples. The table below looks at results from five constructed-response questions that yielded the most zeroes, beginning in 2012 (before Common Core testing) and proceeding from 2013 to 2016 (with Common Core-aligned tests).

You don't need to be a test specialist to know that the purpose of sound testing is not to produce zero-generating questions. The aim is to have students be able to read the passages and to answer them with some

degree of coherence and comprehension.

The pattern here is just the opposite: The questions are getting palpably less intelligible. And when half of students cannot make any sense of a reading passage, we must challenge the measure.

The overall point is that too many questions dumbfounded far too many children. As the table shows, over a quarter of ALL third and fourth graders were frequently left completely lost by numerous questions. When the clock strikes zero, it's time to throw out the clock.

Ever since the New York State grades 3-8 tests were aligned with Common Core, parents and teachers have complained that the tests are not developmentally appropriate, particularly for third and fourth graders. Neither of us is an expert in early childhood or bilingual education and we are not qualified to judge the appropriateness of the reading passages and questions above. But their extreme impact on children must give us pause.

“Science Friction” includes words so difficult that text box definitions are inserted in the story. And “Hattie Big Sky” details life on a farm, using unfamiliar language like “they’ve got some setting left in them,” which is likely outside of the experiences of most New York State test takers.

Remember that third- and fourth-grade students are 8-9 years old.

### Valerie Strauss

Valerie Strauss is an education writer who authors The Answer Sheet blog. She came to The Washington Post as an assistant foreign editor for Asia in 1987 and weekend foreign desk editor after working for Reuters as national security editor and a military/foreign affairs reporter on Capitol Hill. She also previously worked at UPI and the LA Times. Follow 

The Washington Post

Be the first to know.

Try 1 month for ~~\$10~~ \$1

Send me this offer

Already a subscriber? [Sign in](#)