Do Standardized Tests Accurately Measure Student Achievement?

Standardized tests assess whether students have achieved grade-level knowledge and skills, thus determining how well teachers are educating children at grade-level.

Or do they?

The effectiveness of standardized testing has long been debated, with a recent article in Texas Monthly magazine again raising questions about their efficacy.

The story focused on the state’s current standardized test, State of Texas Assessments of Academic Readiness (STAAR), which state law mandates students must pass in grades five, eight and 12 to advance.

Add to this that Texas politicians have floated the idea of determining school funding based on test scores and you’ll understand better the phrase, “high-stakes testing.”

Texas Monthly cited two published analyses of STAAR conducted since 2012 (here and here), both concluding that required reading levels were often higher than student grade levels.

Among other measures, the analyses used the Lexile Framework, which provides grade-level equivalency ratings based on a text’s semantic difficulty (word frequency) and syntactic complexity (sentence length).
The purpose is to match texts to readers’ ability so that they can comprehend around 75 percent of what they read.

Texas Commissioner on Education Mike Morath seemed to dismiss the reports’ findings.

“Morath responded with a lot of jargon and refused to re-evaluate the way the reading test is being administered,” Texas Monthly reported. “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.”

A March 5 hearing of the Texas House of Representatives Committee on Public Education explored these issues.

While several school superintendents expressed concerns about STAAR being above grade level, Morath and others defended the test as accurate (though, admittedly, not perfect) and framed student test achievement (or struggle) as a matter of adequate preparation (or lack thereof).

Also quoted by Texas Monthly was Jeff Cottrill, deputy commissioner of standards and engagement for the Texas Education Agency (TEA), who asserted, “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.’”

That struck me as a throw-away line intended to dismiss concerns out of hand, so I compared a few texts using Lexile’s book search:

- Hop on Pop: 190L (first- to second-grade reading level)
Cottrill is technically correct; yet it feels like a magician’s misdirection to distract the audience from what is really taking place.

Lexile isn’t a flawless system, and educators likely hold different views about its accuracy and validity, but it is a widely used tool for assessing a text’s difficulty that helps ensure students can comprehend most of what they read.

Dismissive comments about Lexile-based analyses are ironic since raw STAAR test results published by TEA include Lexile scores corresponding to student achievement levels.

These appear to confirm outside assessments, as the reported Lexile reading level provided by TEA for students deemed to “master” the STAAR reading tests were higher than grade level for grades four through eight.

Mastering content is defined by TEA as being ready and able “to succeed in the next grade or course with little or no academic intervention.”

As a parent, I would wonder why my child needed to read above grade level to obtain the equivalent of an “A” on their STAAR exam.

A Texas School Alliance report further explains and analyzes the disconnect between the state’s curriculum standards (TEKS) and STAAR tests.
A few closing thoughts and questions, particularly for those making decisions about testing methods and scores:

1. Since the 2001 No Child Left Behind Act, and subsequent legislation like the 2015 Every Student Succeeds Act, the U.S. has increased requirements for public educators.

   They must be “highly qualified” through certification programs and exams, use evidence-based best practices for instruction and obtain continuing education hours for recertification, to name a few requirements.

   If standardized test scores are still not deemed high enough by local, state and national leaders, it seems prudent to look elsewhere for a cause. Perhaps to our assessment measures?

2. Student motivation should be considered. After all, how many K-12 students do you know who like taking tests? How many give their best effort on standardized tests?

   We currently have a system in which teachers and schools are assessed by student performance on tests that many (most?) likely disdain, might not care about how they perform and have little-to-no understanding of the negative consequences for their school or teachers if they don’t perform well.

   Lack of motivation has been cited (see here, here and here) as a contributing factor in lower U.S. student performance on international tests. Could it also be impacting state level assessments?

   For students who give their full effort, receiving a lower-than-expected score (particularly when their teachers have assessed them as being at grade level)
could dampen spirits and lower motivation on future tests.

Standardized testing can be helpful in determining how education systems are functioning only if they are accurate.

This necessitates close and continuous scrutiny through multiple channels – both human and computer-based – to refine exams and ensure they effectively measure student achievement at grade-level.

We must be willing to ask hard questions about our standards of educational assessment, rather than, as too many seem to do, make significant decisions on the assumption that they are valid and accurate measures.

There is too much at stake to do otherwise.