

# NEXT-GENERATION ASSESSMENTS & ACCOUNTABILITY



With support from TASA, John Tanner, executive director of Test Sense and author of “The Pitfalls of Reform,” has written three essays that begin the Texas Accountability Series.

## Why Texas needs a more meaningful school accountability system

by John Tanner

In fall 2017, Texas will join 16 other states in implementing a public school rating system that assigns letter grades to schools and districts. By December 1, 2016, the Texas Education Agency (TEA) had to adopt indicators showing how the A–F ratings will be determined, and by January 1, 2017, TEA must submit a report to the Texas House and Senate Education Committees showing the ratings that schools and districts would have been given if the system had been in place for the 2015–16 school year.

As another Texas Legislature with authority to change the law that established Texas’ A–F system prepares to meet, it is imperative that stakeholders know that the research is clear: A–F school rating systems fail as an indicator of school quality, but there is evidence that supports more meaningful kinds of accountability systems.

With support from TASA, John Tanner, executive director of Test Sense and author of “The Pitfalls of Reform,” has written three essays that begin the Texas Accountability Series. The first essay, “The A–F Accountability Mistake,” provides an overview of A–F systems and their failures. In addition, to question A–F systems is to question test-based accountability, and criticisms of controversial topics are most likely to be heard when solutions accompany the critique, so this essay is accompanied by “Creating a Meaningful Community-Based Accountability System” and “The Misfit Between Testing and Accountability.”

As additional issues related to school accountability arise, the series will be continued to ensure that Texas educators have the information they need to work with policymakers and the public in a meaningful way. Related materials will also be developed and provided to TASA members to assist them in their advocacy efforts.

Following are the executive summaries of all three essays. Educators are strongly encouraged to read the full essays, which are available online at <https://goo.gl/Yrs74f>.

## The A–F accountability mistake

### Executive Summary

**Argument:** The reduction of school quality to a single mark is the purpose of A–F school rating systems. The argument is that a grade will signal a level of quality and make it difficult for low-rated schools to escape scrutiny. Advocates of such rating systems use terms such as “simple,” “clear,” and “transparent” to describe them, and frequently cite competition and subsequent improvement as key outcomes. Former Gov. Jeb Bush of Florida and a number of organizations he supports are the most

vocal proponents of such systems. Florida adopted its system in 1999 and 16 other states have since followed. Texas is scheduled to implement its A–F rating system for the 2017–18 school year.

Research on such systems is surprisingly inadequate given the prevalence of A–F as a policy tool. What does exist is almost universally negative. Florida cites significant gains in the first few years of its program, a fact that is a primary argument in support of such systems. Nevertheless, by Florida’s own admission, the majority of the “gains” were due to changes in the rules, a fact not shared with the Texas Commission on Next-Generation Assessments and Accountability when the Bush-supported organizations offered testimony on this topic in 2016.

Most states with A–F rating systems have adjusted the rules to their systems following implementation so the results more closely match the public and policymakers’ expectations for the distribution of grades. These adjustments call into question the logic behind such systems: It appears they are only declared successful once they reflect a preconceived notion of expectations, not an objective reality.

The few basic rules behind A–F appear simple on the surface but generate an inordinate number of behind-the-scenes calculations and numerous additional rules that render the results unusable for informing change. In many cases schools that perform in a statistically similar manner are awarded vastly different grades, while schools that perform quite differently are awarded similar grades. The reduction to a single grade tends to downplay achievement gaps. In a study of the Oklahoma system, gaps were shown to be wider in higher graded schools than in lower graded schools, and lower graded schools were shown to be performing better with subgroups than higher graded schools.

Based heavily on standardized test scores, A–F school rating systems tend to assign grades in which the socioeconomic status of the school is the single best predictor of the grade, ignoring the efforts being made in some of the most challenged educational environments.

The reduction of a school to a single grade has the tendency to color the judgments and subsequent actions of the entire school, even though each school is a diverse place with the need to serve all students. Reducing a school to a single grade has the predictable effect of telling a school with a good grade that all is well and telling a school with a bad grade that all must change, even though neither can ever be accurate.

**Conclusion:** Rating schools and districts with A–F letter grades is a policy idea that fails every criterion put forth as a reason for having it. It is neither simple nor transparent. It misrepresents a large proportion of what happens in schools by reducing an entire school to a single mark that can only be partially appropriate given the complexity of schooling. In the end, A–F school ratings do more harm than good. They create confusion among educators, and they fail to offer the public useful or accurate information about their schools.

Find the full essay, “The A–F Accountability Mistake,” online at <https://goo.gl/Yrs74f>.

## Creating meaningful community-based accountability systems

### Executive Summary

**Argument:** Organizations consist of people, processes, and systems, all working together toward a defined benefit, frequently known as a mission or purpose. That benefit is the reason for the

organization’s existence. Accountability is the means by which an organization determines its success at providing the benefit, or achieving that mission or purpose. A meaningful accountability system requires each participant in an organization to ask the question: For what am I accountable and to whom?

The state of Texas has answered that question for every educator: accountability is to the state for high or rising test scores. This differs from how educators would answer that same question, given their specific assignments and spheres of influence. Their answers would be deeply influenced by local needs and conditions.

True accountability should be designed such that every educator answers the accountability question and then supplies evidence as to his or her effectiveness. It should support the professionalism of teachers, where most of the educational decisions about children are made. It should guide improvement along the way, rather than offer a post-mortem on a year of schooling derived from a single data collection point at the end of that year.

True accountability is shared. It assigns each task to the appropriate stakeholder with the authority to see it through. It is about improvement. It meets each student wherever he or she happens to be and then moves him or her toward a compelling future. True accountability relies only upon information relevant to the task at hand rather than simply grasping at whatever data or information is available.

**Conclusion:** While it is likely that the state will continue to impose broad-brush judgments on schools in some form or another, schools recognize the disconnect between that broad brush and true accountability for the students placed in their care. True accountability for educators’ decisions is necessarily local,

necessitating the establishment of a meaningful community-based accountability system.

Find the full essay, “Creating Meaningful Community-Based Accountability Systems,” online at <https://goo.gl/Yrs74f>.

## The misfit between testing and accountability

### Executive Summary

**Argument:** Rank order, standardized testing was invented to analyze human traits that could not be readily observed and for which no measuring stick existed. Their invention enables the rank ordering of a population on relative differences, and in turn allows an analysis to proceed in the absence of the measuring stick. The methodology never measured for the amount of anything.

Such tests work by finding a statistical average and then measuring out to the students furthest above and below average to create a ranking. The relative differences between students can then be observed and analyzed, even though a ranking can say nothing of what caused it to come to be.

Because such rankings are based on the aggregate of a student’s experiences with the domain (e.g., numeracy or literacy), the patterns in the rankings will correlate with those experiences. If those experiences have patterns in society, then those patterns will be expressed in the rankings. Given that experiences with numeracy and literacy in the U.S. correlate highly with socioeconomic status, it is not surprising that the rankings do as well.

Ranking is one means by which the patterns in education can be viewed and disrupted. However, rankings are all too often assigned value judgments prior to

knowing the reasons why a ranking is as it is. This is always a mistake: The causes for a ranking need to be determined first. Only then will it become clear if a judgment is warranted, or what types of changes should be supported.

Policymakers noticed that schools they perceived as good had high standardized test scores and declared that all schools should have high standardized test scores. The impossibility of this notwithstanding (e.g., it is impossible for everyone to be above average), policymakers have shown little interest in understanding the realities behind their primary educational accountability instrument. The State of Texas Assessments of Academic Readiness program (STAAR) in Texas is based upon rank order, standardized test methodologies.

**Conclusion:** A methodology designed to show the rank ordering of a population automatically sacrifices any capacity to comment on what caused that ranking. It can serve only as a signal for researchers to begin their search. The quality determination of schools and the passing score for students has, for the duration of education reform, been made via an instrument stripped of any ability to judge quality. This represents a grave concern, as the consequences are extensive.

Find the full essay, “The Misfit Between Testing and Accountability,” online at <https://goo.gl/Yrs74f>.

---

**John Tanner** is the executive director of Test Sense, the author of “The Pitfalls of Reform,” and a consultant to TASA.

