

## A-F Letter Grades: More Questions than Answers

When Representative Jimmie Don Aycock's HB 2804 passed in 2015, [The Texas Tribune](#) reported that, *"The (A-F) approach's supporters argue that it provides a simple and transparent way for parents and community members to understand the performance of their schools."*<sup>1</sup>

In January 2017, TEA released an early example of what A-F letter grades in Domains I-IV might look like for 8,600 public schools and 1,200 districts across the state. This statutorily required "work-in-progress" release allowed local communities and elected officials an opportunity to decide for themselves if an A-F rating system can summarize school performance in a transparent and easily understandable manner.

Judging from coverage in the media, most communities quickly decided that, "It doesn't."

**Let's be clear. An academic accountability system is important to our state and the success of our students. However, it is essential that the system be accurate, credible, and supported by policymakers, educators, parents, and community members alike, for it to be effective in improving students' academic outcomes.**

The provisional A-F system is, at best, a starting point. Per the caveats in Education Commissioner Mike Morath's statement on January 6, 2017 (*"No inferences about official district or campus performance in the 2015–16 school year should be drawn from these ratings, and these ratings should not be considered predictors of future district or campus performance ratings,"* [TEA press releases](#)), the provisional ratings system is of minimal utility. It is not being well-received or supported by those who are interested in improving the Texas public education system, which now serves over 5.3 million students. Questions about system accuracy and credibility abound.

### **The accuracy of A-F letter grades is questionable.<sup>2</sup>**

- The Texas Tribune cited state testing as being among the top education stories of the year (*"[School finance, testing fiasco topped 2016 education news](#)"*), given serious implementation problems that culminated in the state levying a multi-million dollar fine against the vendor. When three out of the five domain calculations are built on these test scores, the accuracy of the letter grades will be in question.
- When high-performing comprehensive high schools — such as those recognized by [U.S. News and World Report](#) and [Newsweek](#) as a top performing high school in the nation, and where almost all of the graduates enroll in selective 4-year universities — receive a "C" for

---

<sup>1</sup> **HB 2804** required the state to change the accountability system for the 2017-18 school year from a 4-Index system (derived primarily from STAAR results and 2 ratings categories of Met Standard and Improvement Requirement) to a 5-Domain system (which includes results from Student Achievement, Student Progress, Closing Performance Gaps, Postsecondary Readiness, Community and Student Engagement) and ratings categories of A, B, C, D, or F.

<sup>2</sup> **Accuracy** is defined as 1: freedom from mistake or error: correctness; and 2a: conformity to truth or to a standard or model: exactness and 2b: degree of conformity of a measure to a standard or a true value: precision. (Webster's Dictionary)

Postsecondary Readiness, the accuracy of the letter grades will be in question. ([“Even high-performing schools get D’s and F’s in Texas’ new grading system”](#))

- A regression model that is applied to only one of four domains does not fully account for the array of challenges faced by public schools. The majority (56.8%) of high-poverty schools earned letter grades of “D” and/or “F” in Domain I (Student Achievement) in the provisional system. When the system fails to take into account factors beyond the schools' control, which are known to correlate with achievement test results, the accuracy of the letter grades will be in question. ([Source: TEA 2015-2016 A-F Ratings: A Report to the 85<sup>th</sup> Texas Legislature, appendix E-65](#))

### The credibility of A-F letter grades is questionable.<sup>3</sup>

- Some problems with the ‘work-in-progress’ A-F system are unique to the methods used by the state to calculate and/or categorize the domain results. For example, a middle school that scored an 89.9 in Domain IV would be given a grade of “F” for that score. If that same school had scored a 99 — a difference of 10.1 points out of 100 — then it would have received an “A.” These cut scores are justified as necessary to differentiate between performance levels; but the resulting grades are neither intuitive, transparent, nor consistently interpreted. When the meaningfulness of the differences between “A’s” and “F’s” cannot be determined, the credibility of the letter grades will be in question. ([Source: TEA 2015-2016 A-F Ratings: A Report to the 85<sup>th</sup> Texas Legislature, appendix B-3](#))
- TEA has published the statement that *“The system should not be built on a forced distribution so that some set percentage of campuses must get an A or an F. [...] it should be mathematically possible that all campuses achieve an A rating”* ([Source: TEA 2015-2016 A-F Ratings: A Report to the 85<sup>th</sup> Texas Legislature, A-F Overview](#)). Unfortunately, espousing the mathematical possibility does not change the fact that in spring 2018, the federal Every Student Succeeds Act, or ESSA, will require the state to identify the lowest performing 5% of Title I schools for comprehensive support and improvement. To do that, TEA must establish cut scores on accountability system data. It doesn’t seem possible to have it both ways. When that happens, the credibility of the letter grades will be in question, perhaps along with the state’s intent.
- Other problems are well-researched and include the relationship between poverty and learning. Research regarding four of the 16 states currently using A-F letter grade accountability systems ([North Carolina](#), [Arizona](#), [Oklahoma](#), and [Florida](#)) consistently reveal that letter grades correlate with the poverty levels of the schools, and **not with performance**. This same poverty / performance relationship is seen in Texas’ “what-if” ratings. When “A” ratings are associated with “Affluence,” the credibility of the letter grades will be in question.

---

<sup>3</sup> **Credibility** is defined as: the quality or power of inspiring belief. (Webster’s Dictionary)

- A-F grades are not productive for school improvement: the ratings alone do not pinpoint what, when, how, or even whether to change what's happening in the classroom. To respond to the ratings in ways that actually help students requires in-depth, local level examination of much more finely grained data. When a letter grade is all that one can see, and there is no immediate ability to understand its cause(s), the credibility of the letter grades will be in question.

The backers of letter grades believe that A-F ratings equate with school quality. They do not. In the 2015 report titled, [“Why School Reports Cards Merit a Failing Grade”](#), the National Education Policy Center found that school A-F grading systems are: 1) not a valid measure of school quality; 2) not valid as a policy instrument to improve school performance; and 3) not a valid measure as a democratic assessment framework.

Upon receiving their campus and district results, superintendents, school boards, educators, parents, and community members around the state publicly questioned A-F letter grades for failure to adequately depict student learning, ignoring the effectiveness of teachers, and not accounting for numerous critical factors that contribute to school performance (such as school funding). These same stakeholder groups have expressed grave concern that the system is inaccurately labeling schools and the children who attend them.

Accountability systems are only useful if their measures are accurate, credible, clear, and useful. Despite good intentions, and a fundamental structure in HB 2804 that is capable of generating a more nuanced understanding of public school performance, the features of TEA's provisional A-F grading system produced ratings that are neither clear nor comparable. The letter grades' lack of clarity and meaning risks leading the public to make unjustified and, in some cases, harmful inferences about differences in the quality of public schools. The 85<sup>th</sup> legislative session provides Texans with one last opportunity to try and answer the many questions about using A-F letter grades in the accountability system before they become permanent in spring 2018.