



ILLUSTRATION / STUART McREATH

Is Test-Based Accountability Dead?

EDUCATION NEXT TALKS WITH MORGAN S. POLIKOFF, JAY P. GREENE, AND KEVIN HUFFMAN

Since the 2001 passage of the No Child Left Behind Act, test-based accountability has been an organizing principle—perhaps *the* organizing principle—of efforts to improve American schools. But lately, accountability has been under fire from many critics, including Common Core opponents and those calling for more multifaceted measures of teacher and school performance. And yet the Every Student Succeeds Act, NCLB’s successor law, still mandates standardized testing of students and requires states to have accountability systems. So: is accountability on the wane, or is it here to stay? If accountability is indeed dying, would its loss be good or bad for students?

In this issue’s forum, we present three different viewpoints on those questions from Morgan S. Polikoff, associate professor of education at the University of Southern California’s Rossier School of Education; Jay P. Greene, professor of education at the University of Arkansas; and Kevin Huffman, former Tennessee commissioner of education.

WHY ACCOUNTABILITY MATTERS, AND WHY IT MUST EVOLVE

by MORGAN S. POLIKOFF



TRY TO THINK of an education policy that 1) has been shown, in dozens of studies across multiple decades, to positively affect student outcomes; 2) has the overwhelming support of parents and voters; 3) reinforces many other policies and facilitates quality research; and 4) has been used widely at the district, state, and national levels for decades or more.

You might be thinking that such a policy doesn’t (*continued on page 52*)

FUTILE ACCOUNTABILITY SYSTEMS SHOULD BE ABANDONED

by JAY P. GREENE



IS TEST-BASED ACCOUNTABILITY “on the wane”? The question is based on a fallacy. For something to be on the wane, it has to exist, and test-based accountability has never truly existed in the United States. Holding people accountable requires that they face significant consequences as a result of their actions. Despite years of “high stakes” student testing, very few of the nation’s 3.14 million public-school teachers have (*continued on page 54*)

IF PARENTS PUSH FOR IT, ACCOUNTABILITY CAN WORK

by KEVIN HUFFMAN



THE GREATEST TRICK the devil ever pulled in education is convincing the American public that we have had test-based accountability. The media and politicians adopted the rhetoric of “high stakes” tests without bothering to ask the question: what, exactly, *are* the stakes? For most adults in education, there were none. Shockingly few public-school educators have lost salary or received a raise or a promotion because of their students’ test (*continued on page 56*)

exist, and if it did, we'd surely want to keep it around. But the truth is precisely the opposite. Such a policy does exist—it's called school accountability—yet the powers that be seem increasingly ready to throw it out and leave education to the whims of the all-but-unregulated free market.

School accountability, specifically test-based accountability, has been a staple of K–12 education policy since the 1990s (and even before that, in some states and districts). Over that time, we've learned quite a lot about it.

First, we've learned that it can work. We've seen this in studies of individual districts, individual states, and the nation as a whole: David Figlio and Susanna Loeb's 2011 review of research summarizes this literature comprehensively. The effects observed in many studies are substantial, especially given that they typically occur schoolwide. The effect of No Child Left Behind (NCLB) on students' mathematics achievement documented by Thomas Dee and Brian Jacob and confirmed by Manyee Wong and colleagues is equivalent to the gain from spending three or four years in an average urban charter school, according to the latest data from Stanford University's Center for Research on Education Outcomes. Accountability doesn't seem to do a great job at closing achievement gaps (though it certainly shines a light on underperformance), but there's considerable evidence that it can raise student achievement.

Second, we've learned that parents and voters feel strongly that accountability is essential. Polls show overwhelming bipartisan support for the common-sense idea that schools receiving public dollars to educate children should be accountable for providing a good education. *Education Next's* 2016 poll reported at least two-thirds support for annual testing among both Republicans and Democrats. In the 2016 PACE/USC Rossier poll of Californians that I led, we asked what schools should be held accountable for; voters rated standardized test results last among the options presented, but 69 percent of them still believed accountability for test results was important. We also know that parents prioritize student achievement when selecting a school for their children. In our increasingly resource-constrained and globally competitive world, this desire for outcomes will only intensify.

Third, we've seen that accountability mutually reinforces other policies and provides essential data to support education research and improvement. For instance, there is suggestive

evidence that charter schools perform better in contexts where accountability is high (that is, where strong authorizing laws shut down poorly performing schools) than where it is weak or nonexistent. Accountability was intended to provide weight to state standards and encourage teachers to implement them, and evidence suggests it does focus teachers' attention on the content that state policymakers want teachers to emphasize.

Not to mention that the data emerging from the same tests used for school accountability have powered a revolution in education research that has allowed scholars to dramatically improve the relevance and rigor of their work.

Finally, we've learned a lot about how to design accountability policy to better target the schools that most need improvement. It is now generally understood that the simplest performance measures—those that defined test-based accountability under NCLB—mainly tell you who's enrolling in a school, not how well the school is educating those

students. We know that performance indexes and growth measures are much fairer and more accurate ways to classify school performance. There's also a growing consensus that in the next generation of accountability policies, we must broaden the criteria beyond test scores, and the new federal education law, the Every Student Succeeds Act (ESSA), encourages this kind of creative rethinking.

We've also learned that the design of accountability policies can affect the way teachers respond to them. For instance, policies that focus attention on raising the achievement of low-performing students may be more effective than those that offer rewards for high student performance in general. And teachers do seem to respond rationally to accountability policies by focusing more on the grades and subjects that are tested. As for concerns about NCLB's negative impact on teacher working conditions, Jason Grissom and his colleagues have shown that the law's implementation did not diminish teachers' job satisfaction or increase their levels of stress. While the unintended consequences of accountability can be pernicious, they can also be addressed, at least in part, through policy design.

Countering the Opposition

Despite this track record of modest success, many parties seem poised to throw the policy overboard and use the guise of “parental choice” or “local control” to return us to a



time when we had little idea which schools were educating children well and which were not. The opposition to accountability in education is largely political; my 2016 analysis of California poll data, for instance, found that disapproval of President Obama was among the strongest predictors of Common Core opposition, and *Education Next* and others have routinely found that voters support “common standards” or “common assessments” when they are not tied to the Common Core name.

There are of course more principled concerns with accountability, and it is worth taking a moment to address them. One issue is that accountability in general, and test-based accountability in particular, can have negative effects on instruction, such as a dumbed-down, narrowed curriculum. This problem can be addressed in large part by improving content standards and the assessments used to gauge student performance. Recent work I co-led with Nancy Doorey indeed finds that the two state-assessment consortia have made considerable improvements over even the best NCLB-era tests. This issue can also be addressed by broadening the set of indicators against which schools are evaluated, which many states are poised to do under the new federal accountability law.

Another concern is that the tests used for accountability do not predict important life outcomes, and thus that we

on choice, especially more extreme forms such as vouchers, has begun to sour). But choice without accountability is unlikely to work. Without test results, for instance, we would not know that online and virtual charters appear to be demonstrably harmful to students, as are many Louisiana private schools attended by students using vouchers. Nor would we know that Boston’s well-regulated charter high schools produce truly stunning positive effects on students’ test scores and early college decisions. Choice programs that do not contain accountability provisions offer us zero assurances that educational dollars are being well spent.

Where should we go from here? We must continue to recognize that the design of accountability policy matters, and we must refine our policies over time. ESSA allows states to do this. It allows states to include better test-based measures of school performance, and they should. It allows them to incorporate measures of school climate, student attendance and discipline, and progress toward college and career readiness, and states should adopt and experiment with these measures. It allows them to target consequences on a smaller subset of low-performing schools and move away from NCLB-era interventions that were largely ineffective, and states appear to be focusing their efforts on more promising interventions that target growth and effective practices. Will the next round of state accountability policies be perfect?

We must continue to recognize that the design of accountability policy matters, and we must refine our policies over time, as the new Every Student Succeeds Act allows states to do.

might be focusing on the wrong things. To be sure, studies do not show a perfect one-to-one relationship between impacts on test scores and impacts on later life outcomes—no one expects they would. But several studies do show longer-term effects of accountability policies; we have strong evidence from Raj Chetty and colleagues that impacts on test scores do predict impacts on other important life outcomes; and, again, many states appear poised to broaden accountability measures beyond just test scores.

What Comes Next?

There is no doubt that the coalition that once supported accountability policy has frayed. The Republican leaders in the executive and legislative branches, which once championed accountability, have turned to school choice as the primary strategy to produce reform (even as public opinion

They will not. Will they be better than what they replaced? They almost certainly will.

Over the last several decades, we have made real, if incremental, progress in education. Test scores on the National Assessment of Educational Progress are up for every student subgroup (even accounting for the downward blip in 2015), and graduation rates are, too. Accountability systems have worked well with other reforms—such as effective choice policies, the expansion of early-childhood-education and other school-readiness programs, and efforts to improve the teaching force through evaluation and tenure reform—to improve education for children around the country. There is simply no reason to think that abandoning accountability at this point would be an effective strategy. The coming years will see new and creative uses of accountability in states and districts. We must encourage and study this innovation if we are to continue improving America’s public schools. ■

GREENE

(CONTINUED FROM
PAGE 51)

ever lost a job, had their pay reduced, or otherwise faced meaningful consequences because of these test results.

It's true that under the No Child Left Behind (NCLB) accountability regime, schools have been given labels, such as "in need of improvement." Some have even been threatened with reorganization or closure. But these threats have only rarely been carried out, and the educators in these schools have typically just been reshuffled to other locations or new management. This enterprise does not constitute true accountability. It's more akin to "double secret probation," the toothless threat imposed on the party-mad frat boys of the 1978 film *Animal House*.

The real question we should address here is whether the hollow threats of this double-secret probation are on the wane. I think they are, and I say good riddance. While testing has failed to produce meaningful accountability, it has distorted the operation of schools to the detriment of educational quality—and it has proven politically unviable.

Educational Harm

Test-based accountability is essentially a central-planning exercise similar to that used by officials in the Soviet Union in attempting to manage the country's economy. In both cases, a distant official selected a particular goal for production, focused on a limited set of metrics to assess whether goals were met, and then threatened to impose rewards or sanctions based on whether those metrics showed desired results. Central planning failed in the Soviet Union, and it is failing here in public education—and for similar reasons.

First, education goals established by distant officials cannot possibly capture the diverse spectrum of local priorities in our nation. Officials have focused on improving math and reading ability, but emphasizing those subjects has come at the expense of other goals. Several studies, including a recent paper by the University of Virginia's Daphna Bassok and colleagues, as well as widespread reports from educators, show that schools have shortchanged history, science, physical education, art, music, and civics. They've also cut back on culturally enriching field trips. Even within math and reading, schools tend to focus narrowly on tested items, which often exclude poetry, literature, and more abstract math.

Providing students with math and reading skills that are

useful in the workplace is a worthy goal of education, but so is helping students become good citizens—cultured, tolerant, self-disciplined, and creative. With test-based accountability, distant officials have imposed their preferences on the rest of us. In addition, studies such as the ongoing research of David Grissmer and colleagues indicate that long-term achievement in math and reading depends on a broader education that includes the type of general knowledge conveyed by history, science, art, and music. Paradoxically, a narrow focus on math and reading may undermine later success in math and reading.

Second, the limited metrics used to assess math and reading achievement are easily gamed and further distort the educational process. If success is defined by the percentage of students who exceed a threshold for proficiency, officials will be tempted to lower the bar for what constitutes "proficient." Schools will also be tempted to focus on students whose performance is below but close to the proficiency threshold, neglecting both high achievers and students who are unlikely to pass even with a reasonable amount of extra attention. School administrators and teachers will be tempted to cheat, as in the recent scandal in Atlanta, or to narrow their instruction, as mentioned earlier. And given that test-based accountability systems are almost entirely built around proficiency levels rather than growth, schools can *appear* more successful if they can avoid serving too many students who are difficult to educate.

Third, schools are gradually figuring out that few real consequences will befall them if they fail to meet the imposed metrics. School leaders' bluffs about mass firings could only be sustained for so long. As a result, scores on the National Assessment of Educational Progress rose in the early years of test-based accountability, but more recently, those gains have stalled.

Furthermore, test-based accountability is built on the assumption that test results are reliable proxies for success later in life, but research has called that assumption into question. It's true that test scores are correlated with some measures of later life success, but for test-based accountability to work we would need to see that *changes* in test scores caused by schools are associated with *changes* in later life success for students. Test-based accountability proponents can point to research by Raj Chetty and colleagues that shows a connection between improvements in test scores and improved outcomes in adulthood, but their work examines testing from the 1980s, prior



to the high-stakes era, and therefore does not capture how the threat of consequences might distort the relationship between test-score changes and later life outcomes.

Furthermore, findings such as theirs are more the exception than the rule. A growing number of studies show a disconnect between short-term progress on test scores and long-term success. That is, even when we measure the extent to which schools contribute to student test-score growth—something that test-based accountability systems rarely do—we cannot consistently predict which programs or schools will help students be more successful later. We cannot centrally plan success if we cannot reliably predict success.

Political Weakness

The educational failures of test-based accountability, as detrimental as they are, will not spell its demise. Rather, accountability that centers on testing is doomed because it has many political adversaries but no enduring political constituency. Parents have never rallied to demand that their children be tested more, that tests be used to retain students or prevent them from graduating, or that tests be used to determine teacher pay or employment. Educators revile test-based accountability even more. Test-based accountability was initiated by policy elites frustrated over rising education costs and subpar results. But elites cannot sustain such a

illustrates the political folly of test-based accountability. Common Core attempted to transform largely symbolic accountability systems into something tougher, which is precisely why it failed. The standards were an effort to better articulate the proper goals of education. The federally subsidized tests aligned to Common Core and developed by the SBAC and PARCC consortia were intended as the rigorous metrics for this stronger accountability regime (see “The Politics of the Common Core Assessments,” *features*, Fall 2016). And centralized teacher-evaluation systems being pioneered by the Gates Foundation in their Measures of Effective Teaching effort were supposed to impose meaningful consequences for failure to perform well on those metrics.

Even these baby steps toward a real accountability system produced a fierce political backlash, led largely by suburban middle- and upper-middle-class families. Such families are accustomed to having significant autonomy with respect to what and how their children are taught, either by choosing the public or private schools their children attend or by influencing locally elected and appointed school officials. By its nature, test-based accountability shifts control away from these parents. Suburban families see Common Core as an infringement on their autonomy, and they have the savvy to fight back. As they do, we are seeing fewer than half of the states sticking to one of the Common Core testing consortia. Soon Common Core will become the same type of nonentity

**Common Core attempted to transform largely symbolic
accountability systems into something tougher, which is precisely why it
failed, illustrating the political folly of test-based accountability.**

policy in the face of opposition from educators and families. American politics is shaped by the activity of organized interests, not poll results. Other countries may be able to impose meaningful systems of test-based accountability, but the decentralized nature of American education and politics gives far more power to organized groups of upper-middle-class families and educators than to the technocratic elite.

The political weakness of test-based accountability helps explain why there are no meaningful consequences attached to it. Opponents have not been able to repeal testing, since there is broad support for *information* on student achievement—even partial and distorted information—but these adversaries have effectively neutered the consequences of accountability. So, the Every Student Succeeds Act continues to require testing, but the accountability piece is even more anemic than it was under NCLB.

The collapse of the Common Core State Standards

it was meant to replace.

What might constitute real accountability in K–12 education? The power of middle- and upper-middle-class families to exercise control over how and what their children are taught is one example. Suburban schools that stray from parental preferences may lose students and revenue or have to answer to angry parents. Test-based systems are politically doomed because middle- and upper-middle-class families tend to prevail in education politics. This phony accountability harms education and undermines schools’ direct accountability to parents. Rather than doubling down on such futile efforts, education reformers should seek to expand true accountability by increasing school choice for more families. The solution to rising costs and subpar results is not central planning but greater control over education on the part of all families, rich and poor. ■

results. By the same token, very few teachers have been counseled onto a different career track or been required to complete targeted professional development. Outside of a handful of states and cities, true test-based accountability has never been implemented.

This is a problem, because the stakes *are* extremely high for students. Parents and teachers complained—with some legitimacy—that the No Child Left Behind (NCLB) era set loose an avalanche of weak fill-in-the-bubble tests to assess student mastery of watered-down state academic standards.

Yet even these substandard assessments were predictive of future life outcomes. As Tennessee's commissioner of education, I could look at the results on the state's old tests and make a highly accurate prediction how any particular 8th grader would eventually score on the ACT, which, itself, is highly predictive of completing a two- or four-year postsecondary degree. Correlation doesn't equal causation, but the test results—with all of their inherent weaknesses—gave a strong indication of how students were faring in our system.

While we spent recent years pretending that a teacher might lose his or her job because of an 8th grader's poor test results, we gave short shrift to the reality facing the 8th grader: a lifetime of truncated opportunities dictated by weak performance at an incredibly young age.

Today, we face two questions. First, will the pseudo-account-

will be public scrutiny and potential embarrassment as a result of receiving a lower "grade" or being placed on a watch list—in most cases, with little formal consequence.

With states now appropriately crafting accountability frameworks that focus not just on test scores but on multiple measures, we also will hear less heated rhetoric about the consequences of poor results. The draft state ESSA plans that I have seen cite measures such as technical assistance for districts and "continuous improvement feedback cycles." Toning down the

rhetoric of accountability—particularly when the realities didn't match the heated language—makes sense, as long as we don't lose our resolve to use student results as a barometer of whether educators are succeeding.



Accountability Works

Because, while teachers and parents may have grown tired of accountability, here's the rub: test-based accountability, even executed poorly, works. From 1999 to 2011, during the heyday of NCLB and its state-level predecessors, overall student scores

improved on the National Assessment of Educational Progress (NAEP) in both reading and math (see Figure 1). Furthermore, scores rose faster for African American and Hispanic students, narrowing achievement gaps. State bubble tests may have been weak, the over-fixation on tests in some schools may have been real, but it is a flat-out fact that kids—particularly low-income and minority kids—got a better education.

The new federal education law ensures that annual tests are here to stay, but it also formalizes a reality that has been true for a while: states decide what accountability for results looks like, and their choices will be shaped by public will.

ability of the last 15 years dial back, stay the same, or be transformed into something real? And second, should we care?

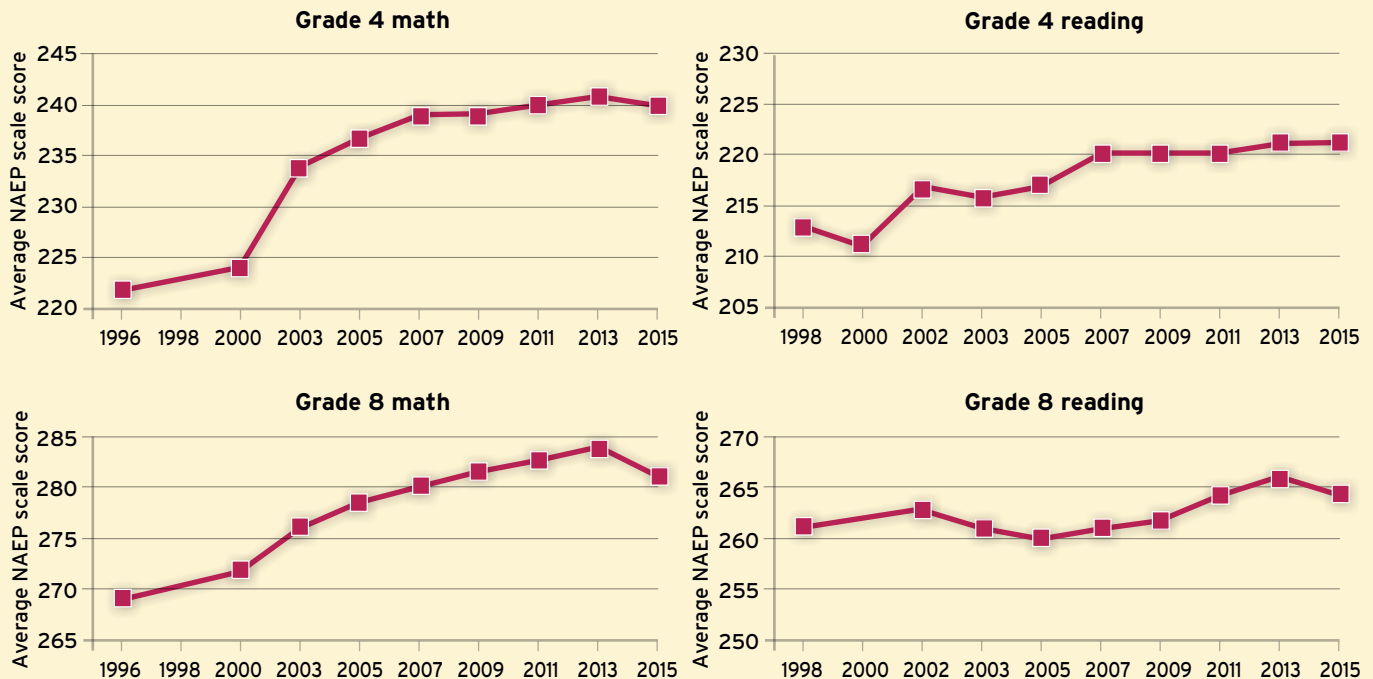
On the first question, my guess is that our attempts at accountability will stay much the same but the rhetoric will dial back. Under the Every Student Succeeds Act (ESSA), states will give annual tests; the results will be published and released; schools will receive some form of rating, based largely on those results; and the very lowest-performing schools will be subject to some form of intervention.

For the average school and school district, the real impact

And while the pace of progress on NAEP has slowed over the past six years, some states and districts continued to make major gains. Tennessee and Washington, D.C., showed the most growth of any state and city in the country, with major improvements in reading and math (and, in Tennessee, also in science). These two places happened to be the leaders in applying test-based accountability to teachers, putting teacher tenure on the line (in Tennessee) and teachers' jobs and salary on the line (in D.C.). While no studies prove what was responsible for the test-score improvements, these results

NAEP Progress under Test-Based Accountability (Figure 1)

From 1999 to 2011, during the heyday of No Child Left Behind and its state-level predecessors, overall scores of public school students improved significantly on the National Assessment of Educational Progress in both reading and math.



SOURCE: U.S. Department of Education, National Center for Education Statistics

would seem to imply that actually holding adults responsible for student progress can have a positive impact on outcomes.

Federalists will often make the case that we should let a thousand flowers bloom, sit back, and wait for the cross-pollination to occur: if states, districts, or schools apply test-based accountability and it works, then others will willingly adopt those best practices.

Reality shows that this is a pipe dream. The national response to improved results in Tennessee and D.C. has been a deafening silence. There have been some laudatory news articles, but precious little cross-pollination. We haven't seen states and districts beating a path to Nashville and D.C. to learn how they improved results. Nor have we seen an increase in states evaluating teachers using student achievement growth and making decisions based on the results. Few if any districts are upending the tenure track and paying teachers different salaries based on student outcomes.

And herein lies the conundrum of accountability and the question of its viability in the coming decade: if test-based accountability works to improve student results but is

unpopular with people who make their living in schools, can we reasonably expect it to find a foothold?

Unfortunately, public school educators are far from the only ones pushing back against hardcore accountability. The fundamental reality is that test-based accountability commands precious little political will all around. While some portray this lack of support as a repudiation of the NCLB era, it actually stems more from the populist policymaking of modern America. Visit any state legislature and you will generally find Democrats spouting union-fed lines about over-testing and demoralized teachers. You will find Republicans repeating talk-radio tropes about the Common Core and its associated tests. And you will find partisans on both sides—given the growing political homogeneity of large cities and rural counties—spouting the merits of local control. That attitude is likely to be reinforced by the new Trump administration and a Republican-controlled Congress.

Right-left fissures in the reform community have reduced the chance of rebuilding a strong bipartisan coalition for

HUFFMAN
 (CONTINUED FROM
 PAGE 57)

accountability. Additionally, the broader education-reform community—the foundations, nonprofits, and think tanks that historically pushed school systems to adapt and change—increasingly have given test-based accountability the cold shoulder. Charter schools have become fetishized at the expense of reforming the traditional public-school system. That's a shame, given that the vast majority of low-income kids today, tomorrow, and 20 years from now are and will be served by traditional school systems.

Do We Care?

We know beyond a shadow of a doubt that some schools, districts, and states are doing better work than their peers. Some are getting better results, and some are driving faster improvements. How do we know this? Because of tests. Because we can see actual evidence that kids have learned things and that schools have improved. Even in the absence of true consequences for low performance, we at least have the ability to identify and learn from the places that are succeeding—if we can spur the necessary actions.

ESSA ensures that annual tests are here to stay, but it also

formalizes a reality that has been true for a while: states decide what accountability for results looks like. Their choices will be shaped by public will. The future of accountability—and of using test scores to improve our schools—will depend on one thing: does the public care enough to advocate for the “eat-your-vegetables,” common-sense annual tests and the associated accountability?

Most parents favor such tests. But if the loudest and most active (read: white upper-middle-class suburban) parents think standardized tests are just an annual annoyance, if these parents and other activist voters choose to disbelieve the results in the fact-free era of modern political discourse, then accountability will be diluted down to the posting of test results and the annual finger wagging of the local news media.

This is where leadership must come into play. It is imperative that governors, state chiefs of education, and other local leaders vocally advocate for the potent change shaper of accountability and convince the public of that power. I am optimistic that state education leaders are availing themselves of the chance to draft stronger, multifaceted measurement systems under ESSA. If voters and parents get behind these systems, and we implement them with fidelity, we will be able to use test results—and other measures—to dramatically improve our public schools. ■


PROCON.ORG

 THE LEADING SOURCE FOR PROS AND CONS
 OF CONTROVERSIAL ISSUES

We research controversial issues and present them in a balanced and primarily pro-con format at no charge.

Dr. Jonathan Haidt calls ProCon.org the "best antidote" to bias

"Among the biggest obstacles to good thinking is what we psychologists call 'the confirmation bias.' It's the tendency to seek out only information that confirms your existing beliefs. ProCon.org is the best antidote to this bias that I have seen. It's not just that it puts disconfirming information right there on the page, where it can't be missed. It's that ProCon.org models open-mindedness, respect for the complexity of truth, and respect for the sincerity of people on both sides of controversial issues. ProCon.org is a boon to our ailing civic culture."


PROCON.ORG


Free



Nonpartisan



Sourced


 Common Core
 Friendly