Holding Accountability Accountable: Taking Stock of the Past 20 Years

By Deven Carlson

May 2021

Key Points

- The No Child Left Behind Act of 2001 (NCLB) had mixed effects on the nation's educational ecosystem.
- NCLB succeeded in shifting the focus from inputs to outcomes, shining a light on performance of different student subgroups, and using reporting requirements to spur development of more-robust education data systems.
- But NCLB included several missteps—namely, setting unrealistic goals and expectations, narrowly focusing on reading and math results, and designing an accountability system with an inflexible one-size-fits-all approach.
- Future accountability legislation should leverage NCLB's high points while charting a path that avoids the law's bumpier portions.

With No Child Left Behind (NCLB) as the entrance ramp, we've now been traveling down the highway of school accountability for two decades. Parts of this road have wound through interesting terrain and scenery, yielding previously unseen views of our education system. Other stretches have been much less pleasant, with bumps, detours, and wrong turns throwing us off course.

Considering the state of the educational world of 2001, it's clear why policymakers across the political spectrum agreed to build this road. The combination of standards, testing, and accountability promised to ensure every student in the United States was proficient in math and reading. But now that we've been on this road for 20 years and still haven't reached this destination—or even one close—it's a good time to stop, look back, and take stock of the productive parts of the journey and the portions we would like to forget.

In surveying the land through which the school accountability highway winds, I identify three major things we ultimately got right about accountability, even if some of these successes were something of a happy accident. I also nail down three things we got wrong, discussing the conditions that led to these missteps and detailing their consequences. In reflecting on the accountability era's successes and failures, I aim to contribute, however slightly, to charting a productive path forward for accountability.

School Accountability's Successes

With today's hyperpolarized politics, it can be difficult to believe there was a time in our not-so-distant past when there was broad bipartisan support for a major piece of federal legislation. But that was exactly the case during NCLB's enactment in January 2002.

At the signing ceremony in Hamilton, Ohio, President George W. Bush touted accountability as the law's driving principle and confidently claimed the legislation would put our nation's schools on "a new path of reform, and a new path of results."[1] Referring to the future of our nation, Sen. Edward Kennedy (D-MA) proclaimed, "No piece of legislation will have a greater impact or influence on that."[2] Reps. John Boehner (R-OH) and George Miller (D-CA), then the chair and ranking member, respectively, of the House Committee on Education and the Workforce, were similarly enthusiastic about the future of our country's public schools. This enthusiasm was even embedded in the legislation, which aimed to make every single student in the country proficient in reading and math by 2014.

Although NCLB didn't produce the educational nirvana politicians promised, the legislation's accountability aspects delivered some very real positives to the country's educational ecosystem. They shifted the focus from inputs to outcomes, shined a light on performance of different student subgroups, and used reporting requirements to spur the development of data systems that have facilitated insights into our education system's operations and effects.

Shifting the Focus from Inputs to Outcomes. The federal government's involvement in public education began in earnest with President Lyndon Johnson's 1965 signing of the Elementary and Secondary Education Act. Spurred by concerns over educational inequity, Title I of the legislation sent money to districts—via state departments of education—to spend on initiatives to improve the quality of education available to disadvantaged students. With only nominal reporting required to access Title I dollars, this initial federal foray into education began a long period in which discussions of dollars dominated education policy debates. Expenditures were about the only aspect of education that were consistently measured, and thus drove policy discussions.

This focus on inputs held sway throughout the 1970s and into the 1980s. In the mid-1980s, though, the notion of educational outcomes began to creep into the policy realm. Most point to the National Commission on Excellence in Education's 1983 report *A Nation at Risk: The Imperative for Educational Reform* as the effective starting point for that discussion.[3] Following its hyperbolic portrayal of the state of American education, the report made some basic outcome-oriented policy recommendations targeted at states, including adopting challenging standards and more-rigorous graduation requirements.

The rest of the 1980s and the 1990s saw a series of initiatives—the National Education Summit, Goals 2000, and the Improving America's Schools Act (IASA)—encouraging states to adopt standards and administer aligned assessments that would, theoretically, provide insight into student outcomes. States were slow on the uptake, though, and at the turn of the century, many states still lacked the infrastructure required to generate meaningful outcome data. For example, Martin Carnoy and Susanna Loeb's comprehensive catalog of states' testing requirements and accountability systems shows that in the 1999–2000 school year, Iowa and Nebraska had no state testing requirements, while other states such as Arkansas, Colorado, and Oklahoma required testing in only one or two grades.[4] And although states such as Florida, North Carolina, and Texas were early adopters of strong school accountability, more than one-fifth of states had no accountability policy at the turn of the century.

NCLB changed that. By threatening to withhold Title I dollars, the legislation effectively ensured that by the 2005–06 school year, states would adopt rigorous standards in reading and math, test students with assessments aligned to those standards, and begin holding schools accountable for the results. In doing so, NCLB flipped the script and pushed student outcomes to the forefront of policy debates. And although the focus on student outcomes hasn't been without drawbacks, on balance it has been a positive development.

Instead of focusing exclusively on inputs—using dollars as a proxy for quality—we are now much more likely to start policy discussions by asking how well schools are serving students. Inputs are clearly part of that discussion, but it has been productive for the policy conversation to progress to viewing inputs as a means to the end of improved student outcomes rather than as ends themselves.

Shining a Light on Different Student Subgroups. Achievement gaps dominate the contemporary education landscape. Everyone with even a passing familiarity of the topography is aware of disparities by race, ethnicity, socioeconomic status, English-learner classification, disability status, and other characteristics. With such broad awareness of these disparities, it's easy to assume that we've always had our finger on the pulse of subgroup performance. That's far from the case, though.

The National Assessment of Educational Progress (NAEP) has disaggregated results for some student subgroups since the early 1990s. However, NAEP lacked universal state participation until 2003; the subgroup disaggregation was done only for the national sample and states that elected to participate. And NAEP's lack of alignment to any common set of state standards further complicated interpretation. IASA also made an effort at disaggregation, requiring states to design their reading and math assessments in a manner that enabled subgroup reporting. But states' blasé approach to compliance with the law—particularly their slow-walking of the required assessments—led to haphazard reporting of results for different student groups.

As was the case in shifting the focus from inputs to outcomes, NCLB was the inflection point for shining a light on subgroup performance. The law required detailed outcome reporting down to the school level for a wide range of student groups including those delineated by race or ethnicity, socioeconomic status, English-learner classification, and disability status. And the legislation's accountability provisions stated that if even a single subgroup of students in a school failed to meet the law's definition of adequate yearly progress (AYP), then the whole school would be classified as failing to make AYP. Such a blunt policy design will have downsides—we'll get to those below—but it directed immediate attention to the performance of different student groups.

Over time, this attention led to a clear-eyed understanding of the states, districts, and schools in which different student groups were and were not scoring well on reading and math tests. This information allowed for a better grasp of the sets of students needing additional attention, more thoughtful consideration of how to help, and the opportunity to detect any indications of success for various improvement efforts. More generally, NCLB's accountability provisions illuminated the outcomes of different student subgroups, which moved us from an "ignorance is bliss" world to one in which we had to consider that our schools served some groups of students much better than it did others. Such illuminations hit particularly hard in states across the Upper Midwest and the Northeast—such as Connecticut, Minnesota, and Wisconsin—that have long prided themselves on the quality of their schools but where subgroup disaggregation revealed some of the largest race-based achievement gaps.

Developing Data Systems. Accountability requires data. NCLB's accountability provisions required data on test scores, attendance, racial and ethnic group identification, sex, free- and reduced-price-lunch eligibility, disability status, and English-learner classification, among other data elements. And the provisions required all this information annually, not just once. At NCLB's enactment, a significant number of districts, and even some states, lacked the technological infrastructure required to collect, maintain, and report these data.

Given the Bill Clinton administration's lax enforcement of IASA's testing and reporting provisions, states seemed to initially expect the Bush administration to take a similar approach with NCLB, considering the law's aggressive timeline as a mere suggestion. But it quickly became apparent that wouldn't be the case, with Secretary of Education Rod Paige making clear that waivers were off the table. This spurred states to kick implementation efforts into high gear, and states generally had their testing and accountability systems up and running by the 2005–06 school year—the original timeline specified by NCLB.

States weren't left completely high and dry in developing their data systems. Just after the passage of NCLB, President Bush signed the Educational Technical Assistance Act of 2002, which authorized the Statewide Longitudinal Data Systems (SLDS) Grant Program. SLDS grants provided states with up to $20 million per award—states could apply multiple times—to help establish their data systems. States varied in how they leveraged these dollars, with states such as California and Oklahoma doing the bare minimum and others such as Florida, North Carolina, and Ohio developing incredibly impressive data systems that link K–12 records with information on postsecondary education, the labor market, public assistance programs, the criminal justice system, and other societal institutions.

These data systems have yielded a stunning number of insights into our education system. We have a detailed understanding of achievement gaps—and disparities in other outcomes—across a range of demographic groups. We've come to more fully appreciate the importance of teachers and how their impact persists long after students leave their classrooms. We know so much about the effects of accountability policies, intended and otherwise. We can detail the ins and outs of different school choice programs such as charter schools, vouchers, and open enrollment. We've painted nuanced portraits of student mobility and attendance patterns. In short, we've accumulated an incredible amount of knowledge about our nation's public education system. And it all started with states working to meet NCLB's testing and accountability provisions.

NCLB's Benefits. Although accountability didn't propel our education system to the heights that politicians promised, it delivered real benefits. Our policy discussions now revolve around student outcomes rather than educational inputs. Data systems have provided a basis for learning so much about our education system, including the performance of different student subgroups.

> Although accountability didn't propel our education system to the heights that politicians promised, it delivered real benefits.

Interestingly, with the partial exception of subgroup reporting, the accountability movement's positive legacies went largely unmentioned at its outset. Politicians didn't tout the data systems that the law would create. They didn't hype the shift in policy discourse. No, they told us that holding our schools accountable would increase their quality to such a degree that all students—every single one—would be proficient in reading and math in just more than a decade. Juxtaposing the rhetoric of accountability with its reality hints at what the federal government might be best suited to do in this realm.

Where We Went Wrong

Accountability clearly failed to deliver on the high-flying promises politicians made in the halcyon days of the early 2000s. But why? What went wrong? Twenty years down the road, we can address these questions.

Three major mistakes were made at the outset of the accountability movement: setting unrealistic goals and expectations, narrowly focusing on reading and math results, and designing NCLB's accountability system with an inflexible, one-size-fits-all approach. In hindsight, we can see how these three early missteps conspired to render accountability, or at least NCLB's version of it, dead in the water.

Setting Unrealistic Goals and Expectations. It's difficult to overstate the degree of rhetorical excess politicians employed when arguing for accountability. President Bush said strong accountability would "make sure not one single child gets left behind in America."[5] Sen. Jon Corzine (D-NJ) professed optimism that strong accountability measures "will help narrow the educational achievement gaps that threaten every child's access to the American dream."[6] At the time, politicians were likely aware of their hyperbole but thought it benign.

Looking back, it's conceivable that NCLB could have survived politicians' aspirational rhetoric and impossible promises. Even in 2001, the public expected a degree of exaggeration and bombast from their elected representatives. But baking the unrealistic goals and expectations into the law—even going so far as to design the entire accountability system around them—was bad policymaking and is ultimately what ruined NCLB's public support.

Setting a goal of universal reading and math proficiency by 2014 effectively ensured we would end up judging accountability, and NCLB more broadly, as a failure. Even at the time, it was clear we were never going to achieve that goal or come close to it.[7] And because the law required states to specify a trajectory toward universal proficiency, the failure judgments wouldn't be put off until 2014. They would begin to trickle in much earlier.

Secretary of Education Arne Duncan told Congress in spring 2011 that without waivers, more than 80 percent of schools would fail AYP that year. And as they did, they provided detractors with ammunition to chip away at accountability.[8] And chip away they did. The annual PDK/Gallup poll on education found that the percentage of Americans viewing NCLB unfavorably nearly doubled between 2005 and 2010, growing from 27 to 46 percent.[9]

The primary danger of using an unattainable goal as the basis for stamping failing labels on a broad range of schools stems from the disconnect between those labels and parents' evaluations of their children's schools. Polling routinely shows majorities of parents grade their children's schools as an A or a B, judgments typically formed by up-close observation of their children's daily experiences.[10] So when a policy comes along and says that their children's school is failing, most parents will trust their own judgment more than an opaquely generated label. And as these discrepancies accumulate, it will ultimately be the policy's validity that parents call into question, not their judgment.

Going forward, policymakers should specify incremental, achievable goals for their school improvement efforts. Of course, goals like this don't provide sound bites like promises of universal proficiency do. But reality doesn't always fit into nice, neat sound bites.

Narrowly Focusing on Reading and Math Test Results. We all want kids to read well and do math at a high level. And we should, as those skills are instrumental for navigating the world. But education is much more than that. It should prepare children for membership in a democratic polity. It should develop social and emotional skills that will be useful in all aspects of life. It should consistently expose students to a range of ideas, subjects, and extracurriculars so they can figure out what draws them in and what repels them. NCLB's accountability system didn't explicitly say education shouldn't do these things, but by basing its ratings almost entirely on students' performance on math and reading tests, it signaled what it valued.

Educators and administrators responded to these signals, going to great lengths to boost math and reading scores in their schools. Many schools devoted full weeks to test preparation strategies leading up to the annual state assessments. Districts often increased instructional time in reading and math at the expense of non-tested subjects such as art, music, and social studies.[11] In the most egregious cases, educators cheated on the tests to satisfy the seemingly bottomless appetite of NCLB's accountability system.[12]

In short, NCLB's accountability system threw the educational ecosystem out of balance and incentivized counterproductive behaviors to game the system. And, adding insult to injury, all the time, effort, and dollars devoted to chasing universal proficiency didn't have much effect. Achievement levels today are only slightly higher than they were 20 years ago. We still see enormous gaps between advantaged students and their less advantaged peers. Well-designed accountability systems shouldn't generate these sorts of outcomes.

The accountability movement's excesses are perhaps best illustrated by the nature of the follow-on reform du jour, social and emotional learning (SEL). SEL explicitly based its appeal on educating the whole child, a not-so-subtle jab at the intense reading and math focus that characterized the accountability era. And the enthusiasm with which folks across the educational landscape—liberals and conservatives, parents and teachers, funders and fundees—responded to this appeal illustrates how eager stakeholders were to move on from a policy conversation dominated by test-based accountability.

Federal Control Without Flexibility. Every consequential aspect of NCLB's accountability system—testing requirements, AYP's definition, the series of sanctions—was designed in DC and dictated to the states, which were given no meaningful freedom to adapt the system to their realities. And while this rigidity perhaps represented a rational response to states' past foot-dragging, it resulted in a one-size-fits-all accountability system that was stunningly ill-suited to handle the variation that is a defining feature of American education.

Further, the NCLB accountability system's inflexible and punitive nature highlights its motivating assumption: Educators weren't trying hard enough, and threatening to punish their schools would make them work harder and increase student learning. It's a mistake to make policy based on assumptions that question educators' motives and efforts. And restricting states and districts from adapting policy in a manner that moves away from such assumptions just compounded that mistake.

A fascinating thought experiment involves contrasting NCLB's accountability system with what might have emerged without such strong federal intervention. At the time of NCLB's 2001 signing, most states had some form of accountability system in place, but there was substantial variation in their designs. And it's interesting to think what might have evolved had states been left on their own. Sure, some states wouldn't have done anything meaningful around accountability, but others were seemingly on track to develop systems that would have avoided many of NCLB's missteps. Systems in states such as Kentucky, Maryland, New Mexico, and North Carolina used test scores as the basis for providing monetary rewards. Other systems, including those in Connecticut and Missouri, used test results as an information point or indicator that particular schools or districts might need additional scrutiny.

Going forward, policymakers should continue the march back toward increased state control of school accountability systems.

NCLB immediately squashed this variation, instead effectively taking the harshest accountability system in the country, Texas', and immediately implementing it everywhere. Given the direction accountability is headed in the post-NCLB years—less punitive, with greater state control—it is instructive to look back and see that some states were ahead of their time. It's also natural to ask whether we might have reached an accountability equilibrium more quickly without NCLB.

Going forward, policymakers should continue the march back toward increased state control of school accountability systems. In my view, the endgame is a set of systems in which states use different indicators—including test scores, graduation rates, and absenteeism—to judge whether a school or system warrants a closer look. These deeper dives, which could be operated out of an office similar to a state auditor, could work to identify the root causes of the concerning numbers. In something as complex as schooling, high-level numbers can tell you only so much, and it often takes setting foot inside a school to get a sense of what's going well and what isn't and to make productive recommendations about how the school might go about any improvement process.

The Cost of These Missteps. These three major missteps helped turn the tide against accountability, putting a bad taste in the mouth of the public and leading to the scaled-down version in the Every Student Succeeds Act. Such a retreat could be healthy, giving accountability proponents time to reflect on what went wrong the first time and to work on developing more thoughtful and productive approaches to ensuring school quality. But the long-term cost of these mistakes might be the prohibition of any idea with even a whiff of enhanced accountability from the policy arena. It's possible that these mistakes so poisoned the well that any accountability-oriented reforms are off the table for the foreseeable future.

Conclusion

The accountability movement began with the best intentions, promising to pave a road to an educational paradise where every single student in the country was proficient in reading and math. We've been on this road for 20 years and never arrived at this destination, instead ending up somewhere that doesn't look dramatically different from where we started.

The journey has certainly been interesting, though. Along the way, we've collected souvenirs we will long value—a focus on outcomes, impressive data systems, and a nuanced understanding of the outcomes of different student subgroups. But some wrong turns and detours sent us down desolate roads we'd like to forget. As we plan our next trip, we'd do well to leverage the high points of the last journey and chart a path that avoids the bumpier portions.

About the Author

Deven Carlson is an associate professor of political science and presidential research professor at the University of Oklahoma. He has written extensively on accountability, school choice, school closures, and the politics of education.

Notes

1. White House, "President Signs Landmark No Child Left Behind Education Bill," press release, January 8, 2002, https://georgewbush-whitehouse.archives.gov/news/releases/2002/01/20020108-1.html.
2. 147 Cong. Rec. S13419 (daily ed. December 18, 2001), https://www.congress.gov/crec/2001/12/18/CREC-2001-12-18-senate.pdf.
3. National Commission on Excellence in Education, *A Nation at Risk: The Imperative for Educational Reform*, US Department of Education, April 1983, https://www2.ed.gov/pubs/NatAtRisk/index.html.
4. Martin Carnoy and Susanna Loeb, "Does External Accountability Affect Student Outcomes? A Cross-State Analysis," *Educational Evaluation and Policy Analysis* 24, no. 4 (Winter 2002): 305–31, https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.1075.1880&rep=rep1&type=pdf#:~:text=Susanna%20Loeb&text.
5. White House, "President Bush Discusses Education Reform," press release, July 5, 2001, https://georgewbush-whitehouse.archives.gov/news/releases/2001/07/20010705-6.html.
6. 147 Cong. Rec. 26,599 (2001), https://www.govinfo.gov/content/pkg/CRECB-2001-pt19/html/CRECB-2001-pt19-Pg26577-8.htm.
7. Thomas J. Kane and Douglas O. Staiger, "Rigid Rules Will Damage Schools," *New York Times*, August 13, 2001, https://www.nytimes.com/2001/08/13/opinion/rigid-rules-will-damage-schools.html.
8. Deven Carlson, "Testing and Accountability: What Have We Learned and Where Do We Go?," in *Bush-Obama School Reform: Lessons Learned*, ed. Frederick M. Hess and Michael Q. McShane (Cambridge, MA: Harvard Education Press, 2018).
9. Carlson, "Testing and Accountability."
10. *Education Next*, "2019 EdNext Poll Interactive," August 19, 2019, https://www.educationnext.org/2019-ednext-poll-interactive/.
11. A study conducted by the RAND Corporation drew a representative sample of districts in California, Georgia, and Pennsylvania and surveyed superintendents, principals, and teachers in those districts between 2003 and 2005. About half of principals surveyed reported that they required their teachers to increase time spent on reading and math instruction, with corresponding declines in non-tested subjects. Among teachers, more than 20 percent of Pennsylvania and California teachers reported spending less time on science and social studies instruction after NCLB enactment. About a quarter of California teachers reported a decline in instructional time for arts, music, and physical education. See Laura S. Hamilton et al., *Standards-Based Accountability Under No Child Left Behind: Experiences of Teachers and Administrators in Three States* (Santa Monica, CA: RAND Corporation, 2007), https://www.rand.org/pubs/monographs/MG589.html.
12. The cheating scandal in Atlanta, Georgia, is likely the highest-profile example of the genre, but investigations uncovered evidence of cheating in many school districts across the country, including Columbus, Ohio; Dallas, Texas; and Washington, DC, to name just a few. See Rachel Aviv, "Wrong Answer," *New Yorker*, July 14, 2014, https://www.newyorker.com/magazine/2014/07/21/wrong-answer; Stephanie Banchero, "Columbus, Ohio, School District Hit by Cheating Allegations," *Wall Street Journal*, January 28, 2014, https://www.wsj.com/articles/SB10001424052702303277704579349202004452142; Greg Toppo, "Memo Warns of Rampant Cheating in D.C. Public Schools," *USA Today*, April 11, 2013, https://www.usatoday.com/story/news/nation/2013/04/11/memo-washington-dc-schools-cheating/2074473/; and Jeffrey Weiss and Matthew Haag, "STAAR Scores, Rating Plunged at Top Elementary After Dallas ISD Determined Teachers Had Been Cheating," *Dallas Morning News*, August 14, 2014, https://www.dallasnews.com/news/education/2014/08/14/staar-scores-rating-plunged-at-top-elementary-after-dallas-isd-determined-teachers-had-been-cheating/.

© 2021 by the American Enterprise Institute for Public Policy Research. All rights reserved.

The American Enterprise Institute (AEI) is a nonpartisan, nonprofit, 501(c)(3) educational organization and does not take institutional positions on any issues. The views expressed here are those of the author(s).