Asking too much of accountability: The predictable failure of No Child Left Behind

With the No Child Left Behind Act of 2002, the federal government sought to create a national system of test-based accountability that would ensure that virtually all American students could read and do math on grade level within 12 years.¹ By any reasonable definition, this effort failed. It failed on its own terms, with any gains in student achievement the law generated falling far short of its utopian goals.² It also failed politically, with one of its leading Congressional architects admitting by 2006 that NCLB had become “a tainted brand.”³

That’s not to say that NCLB did no good whatsoever. The law’s requirements that students be tested annually in math and reading in grades 3-8 and once in high school, and that the results be published by school and disaggregated by subgroup, provided parents, teachers, and other citizens with vital information about students’ performance in these foundational subjects. This information shed new light on achievement gaps along lines of race, ethnicity, and class, both across entire states and within specific schools. Rigorous studies confirm that, by requiring states that had not previously adopted test-based accountability systems to do so, the law produced modest gains in student achievement, concentrated in math, that were largest for low-achieving students.³

Nor is NCLB responsible for all of the ills for which it is routinely blamed. There is little evidence, for example, that the law’s accountability requirements increased test anxiety among students or reduced their enjoyment of learning, either overall or in schools at risk of being identified as low-performing.⁴ Though widely perceived as taking resources away from schools serving disadvantaged students, in fact the law led state and local governments to spend more on all of their schools, including those in high-poverty communities.⁵
Yet the benefits NCLB produced do need to be weighed against a raft of unintended consequences. The law’s exclusive reliance on math and reading test scores to gauge school quality contributed to a narrowing of the curriculum in American elementary and middle schools, with teachers devoting considerably more instructional time to math and especially reading at the expense of science and social studies. While arguably desirable as a strategy to enhance literacy and numeracy, the law’s muted effects on student achievement suggest that this massive reallocation of educators’ efforts has been counter-productive. In too many schools, educators responded to the pressure to boost scores by resorting to crude methods of test prep or even to outright cheating, producing increments in measured performance without enhancing student learning. Harder to quantify, but likely substantial, was the impact of the law’s one-size-fits-all-states approach on the pace of innovation in the design of educational accountability systems and other school reform efforts.

Whether the unintended consequences of NCLB outweighed its benefits is impossible to determine with any certainty. That judgment would require evidence on how the law influenced a range of outcomes we struggle even to measure, much less to weigh against those we can. Fortunately, however, a final reckoning is needed neither to deem NCLB a failure nor to derive lessons from its demise.

In this chapter, I will trace the roots of NCLB in prior federal and state reform efforts and explain what made it distinct: tight prescription with respect to the design of school accountability systems coupled with a newfound commitment on the part of the U.S. Department of Education to ensure that states complied with the law’s mandates. Contrary to conventional wisdom, the accountability systems the law forced states to adopt were hardly punitive. Poor test results did not threaten educators’ jobs and carried no stakes at all for students. But the law was
quite demanding of states, which risked the loss of federal funds if they did not go through the motions of complying with its directives.

“Go through the motions” is exactly what they did—and what they permitted on the part of local school districts under their control. Left to set their own standards for student proficiency, most states set undemanding benchmarks. Told to give students in schools where student proficiency rates lagged statewide targets the choice of a better alternative, school districts did little to notify parents of their options. Asked to intervene in persistently low-performing schools, most pursued only superficial measures. Few, if any, states and districts offered schools greater flexibility in the use of resources even as they subjected them to ever-higher performance targets.⁹

A first lesson to be drawn from the national experience under NCLB thus concerns the federal role in American education. As a ten-percent investor in a roughly $700 billion public education system, the federal government lacks the capacity and license to intervene directly in local schools; it must inevitably rely on states and districts to act on its behalf.¹⁰ NCLB’s implementation confirms that the threat of lost funding is enough to drive superficial compliance with even far-reaching federal directives. It is often insufficient, however, to compel local officials and educators to do so in a manner consistent with Congressional intent. This recognition should counsel humility among federal policymakers about their ability to advance reform.

NCLB’s failure is also instructive as to the limits of test-based accountability as a strategy for improving the performance of schools and school systems, regardless of whether it is pursued at the state or federal level. For more than three decades, elected officials across the political spectrum have advocated the development of content standards in core academic
subjects and the regular administration of aligned tests in a well-intentioned effort to enhance system coherence and ensure that the resources invested in public education produce greater opportunities for American students. In seeking to measure the outcomes produced by schools and the educators within them, these efforts amount to nothing less than an attempt to alter the nature of schools as organizations—and therefore the management strategies available to improve them.\(^{11}\)

In his classic work on *Bureaucracy*, political scientist James Q. Wilson used schools as the prototypical example of a “coping organization”—that is, an organization in which “effective management is almost impossible” because managers are able to observe neither the outputs (i.e., work) of the organization’s employees nor the outcomes they produce.\(^{12}\) “A school administrator,” he wrote, “cannot watch teachers teach (except through classroom visits that momentarily may change the teacher’s behavior) and cannot tell how much students have learned (except by standardized tests that do not clearly differentiate between what the teacher has imparted and what the student has acquired otherwise).”\(^{13}\)

Wilson contrasted schools with “production organizations” like the Internal Revenue Service and the United States Postal Service, where both outputs and outcomes are visible and “managers have an opportunity…to design a compliance system to produce an efficient outcome.”\(^{14}\) Whether they will do so hinges on whether the political process that structures the organization’s work prioritizes efficiency as opposed to, say, the satisfaction or job security or its employees. Yet the prospect of performance-oriented management was understandably attractive to reformers frustrated with the achievement of American students relative to their peers abroad and with the persistence of gaps in achievement along lines of race and class. With NCLB, the federal government sought to convert schools nationwide into production organizations.
Consistent with research on its state-level predecessors, the implementation of NCLB confirmed that educators do respond to the incentives created by school accountability systems, even when the performance measures underpinning those systems are crude and the stakes attached to them weak. This might suggest that any distortions the law produced can be solved through better design. It is the case that the most salient flaws in the measurement system that NCLB required states to adopt—its unreasonable performance targets, its emphasis on student proficiency rates, its attention to the level at which students achieve rather than how much they learn over time—were evident when the law was passed and became all the more so over time. Moreover, those flaws have solutions that, from a design perspective, are straightforward to implement. If the incentives created by accountability systems work, perhaps we just need to strengthen them, apply them more broadly (to individual teachers, for example, in addition to schools), and be sure to get them right.

However alluring, this interpretation faces technical and political challenges. The technical challenges stem from the fact that, although standardized test scores provide one important indicator of the extent to which schools have equipped students to succeed in post-secondary education and beyond, there is much that they miss. Common sense tells us (and mounting evidence confirms) that a range of skills, competencies, and mindsets beyond literacy and numeracy are at least as important in shaping students’ life prospects—and that schools can help foster their development. Yet efforts to develop valid and reliable measures of those skills remain in infancy. In part for that reason, we lack a solid understanding of whether and when efforts to raise academic achievement complement or detract from efforts to develop students more holistically. What is already clear, however, is that schools and the educators within them
vary in their effectiveness in improving non-test outcomes that matter for students’ academic and labor-market success and that this variation is only weakly related to their effects on test scores.\textsuperscript{18}

This evidence has important implications for test-based accountability. In complex production processes with multiple goals, incentives to improve a subset of those goals risk distracting attention from others. When goals that are not incentivized also go unmeasured, it is impossible to weigh what is lost against what is gained as a result. This dynamic accounts for the reallocation of instructional time across subjects under NCLB and its state-level predecessors; what was tested became what was taught. Even if broadened to encompass achievement in subjects beyond math and reading, accountability systems based on test scores alone are likely to discourage educators from pursuing other valued goals. At a minimum, they will provide only a partial measure of educators’ effectiveness in preparing students to succeed.\textsuperscript{19}

The political challenges confronting test-based accountability systems reinforce the technical challenges. A premise of outcome-based accountability policies of any kind is that external pressure to improve measured performance, whether through explicit rewards and sanctions or simply through increased transparency, will drive behavior change—that it will lead educators to work harder or differently, or lead administrators to make politically unpopular decisions. As this occurs, however, the outcome measures on which accountability systems rely are likely to come under attack from organized interests who benefit from the status quo. Meanwhile, defenders of the measures will be scarce: while most Americans favor educational accountability policies in the abstract, that support is diffuse. Given this imbalance of interests, not to mention the imperfections of even the best educational measurement systems, attacks on accountability metrics are apt to succeed.\textsuperscript{20}
This dynamic was evident in the education sector’s response to NCLB from the outset and contributed to the sharp decline in the law’s popularity over time. Perhaps the best illustration of its significance, however, came in response not to NCLB itself, but rather to the Obama administration’s effort to extend its logic to individual teachers through its Elementary and Secondary Education Act (ESEA) Flexibility Program. As NCLB’s deadline for universal proficiency approached, its accountability system became increasingly unworkable, with a majority of schools in many states identified as low-performing. With Congress unable to agree on a replacement, the Obama administration in 2011 took matters into its own hands, offering limited flexibility from NCLB requirements to states willing to pursue a set of strategies it identified. These strategies included test-based evaluations of teachers and school leaders and more aggressive efforts to intervene in their lowest-performing schools.

In asking states to overhaul their approach to educator evaluation, the administration sought to be sensitive to the limitations of test-based performance measures. While it required that they be given substantial weight, for example, it directed states at a minimum to complement them with rubric-based observations designed to provide a more holistic appraisal of teachers’ classroom practice. Yet these steps were hardly sufficient to assuage the concerns of teachers unions opposed in principle to tying high-stakes personnel decisions to evaluation results. This opposition centered above all on the validity of value-added measures of teacher performance and the tests used to generate them. In some states, the unions encouraged and lent organizational heft to a burgeoning movement among parents to opt their children out of state tests. By 2015, members of Congress were inundated with constituent demands to do something about the perceived scourge of over-testing. It seemed possible that lawmakers would do away with NCLB’s core annual testing requirement altogether.
They did not, averting what in my view would have been a final, and perhaps most significant, failure of NCLB. The Every Student Succeeds Act continues NCLB’s requirement that students be tested annually in reading and math while returning to states most decisions about how to identify low-performing schools and the steps taken to improve them. In doing so, it ensures that citizens continue to have basic transparency about schools’ success in promoting basic literacy and numeracy, skills that are central to core mission of public education and prerequisites for many other educational goals. Yet it enables states to complement that information with additional indicators of school quality or student success and, as important, to set reasonable expectations with respect to the pace of improvement.

In short, the information provided by test-based accountability systems is an essential component of education reform efforts. Those systems cannot, however, be relied upon as the primary drivers of improvement in American education. Reformers who ask too much of test-based accountability are bound to be disappointed.
References


---

1 Peterson & West, 2003.
3 Dee & Jacob, 2011; Wong et al., 2015.
4 Reback et al., 2014; Dee et al., 2013.
5 Dee et al., 2013.
6 West, 2007; Dee et al., 2013.
7 See, e.g., Neal & Schanzenbach, 2010.
8 West, 2017.
9 Manna, 2010; Peterson, 2016.
11 Elmore et al., 1996.
12 Wilson, 1989, p. 175.
14 Ibid, pp. 159-60.
Note that this analysis actually understates the problem facing test-based accountability systems, in that it presumes that there is in theory a correct set of weights that schools should assign to each educational goal. To the extent that citizens vary in their preferences over the set of outcomes schools should prioritize, and policymakers want to respect those preferences, the situation is even more complex.