

# ARTICLE

## WHEN PERFORMANCE MATTERS: THE PAST, PRESENT, AND FUTURE OF CONSEQUENTIAL ACCOUNTABILITY IN PUBLIC EDUCATION

SANDY KRESS\*, STEPHANIE ZECHMANN\*\*,  
& J. MATTHEW SCHMITTEN\*\*\*

*The “consequential accountability” model of education reform involves creating explicit standards for students, testing against those standards, and assigning consequences to schools for failure to meet those standards. The No Child Left Behind Act of 2001 (“NCLB”) represents the current high watermark of the consequential accountability movement and of federal involvement in education reform. NCLB conditions the receipt of Title I federal education funds on schools implementing consequential accountability systems. The Act is currently up for reauthorization, which presents a valuable opportunity to use lessons learned from past consequential accountability experiments to improve the legislation by strengthening federal accountability policies. Toward that end, this Article reviews the evolution of consequential accountability in the United States and argues that this reform model has led to significant gains in student achievement vis-à-vis students at schools without consequential accountability. The Article evaluates different state consequential accountability models and contrasts them with the federal model embodied in NCLB. Finally, the Article offers suggestions for which aspects of NCLB should be maintained, and what improvements would help maximize the effectiveness of consequential accountability in bettering America’s system of public education.*

*People shouldn’t fear accountability. They ought to welcome an accountability system as a useful tool to make sure no child is left behind, a useful tool to make sure they make constructive improvements in their school districts and in their schools.*

*—President George W. Bush<sup>1</sup>*

### I. INTRODUCTION

“Consequential accountability” can be broadly defined as any education accountability mechanism that has the following three elements: (1) ex-

---

\* A.B., University of California at Berkeley, 1971; J.D., University of Texas Law School, 1975; Former Senior Education Adviser to President George W. Bush, current Fellow at the George W. Bush Institute, and Senior Counsel, Akin Gump Strauss Hauer & Feld LLP.

\*\* B.A., American University, 2009; J.D. Candidate, Harvard Law School, Class of 2012.

\*\*\* B.A., American University, 2009; J.D. Candidate, Columbia Law School, Class of 2013.

<sup>1</sup> George W. Bush, Remarks on Implementation of the No Child Left Behind Act of 2001 (Sept. 4, 2002) in THE AMERICAN PRESIDENCY PROJECT (John Wooley & Gerhard Peters eds.), <http://www.presidency.ucsb.edu/ws/index.php?pid=73078> (last visited Oct. 26, 2010).

PLICIT, publicized standards; (2) regular testing against those standards; and (3) consequences linked to performance.<sup>2</sup> In 2001, the No Child Left Behind Act (“NCLB”)<sup>3</sup> dramatically changed the federal education landscape by requiring states to develop and implement “consequential accountability” systems as a condition of receiving Title I federal education funds, which the federal government provides to schools and districts with a high percentage of students from low-income families.<sup>4</sup> While NCLB brought consequential accountability into the national spotlight, it is only the most recent chapter in a long history of reformers advocating for greater accountability to improve the American public school system. The consequential accountability movement itself began in the 1990s, when nearly forty states integrated the concept into their education policies and when Congress introduced it into federal education policy through the Improving America’s Schools Act of 1994 (“IASA”).<sup>5</sup>

Studies have shown that the implementation of consequential accountability systems in the 1990s likely resulted in significant gains in student achievement in the states that employed them. The momentum created by the early success of the accountability movement provided much of the impetus for the passage of NCLB in 2001. Now, nearly ten years after the law’s passage, substantial additional data and research are available to inform future legislation and policymaking.

This Article will discuss the past, present, and future of consequential accountability in the United States and offer evidence to support the proposition that consequential accountability has led to gains in student achievement in the late 1990s and 2000s. In Part II, the Article presents a selective history of education reform in the United States, highlighting the events that played an important role in the development and evolution of consequential accountability in the state and federal context.

In Part III, the Article evaluates the empirical evidence concerning the effectiveness of consequential accountability systems. First, it reviews the literature on the use and effect of consequential accountability in the 1990s, which makes the case that states employing consequential accountability models saw statistically significant gains in student achievement during the 1990s as measured against their “non-accountable” counterparts. Next, it

---

<sup>2</sup> The definition of “consequential accountability” used in this Article is essentially the same definition provided in Eric A. Hanushek & Margaret E. Raymond, *Does School Accountability Lead to Improved Student Performance?*, 24 J. POL’Y ANALYSIS AND MGMT. 297, 311-12 (2005) (differentiating between states with just tests and standards—or “report card” states—and states that attach consequences to school performance—or “consequential accountability” states). Hanushek and Raymond define a state as a “consequential accountability” state if it both reports the results of standardized testing and attaches consequences to school performance vis-à-vis the test results. *Id.* at 306.

<sup>3</sup> No Child Left Behind Act of 2001, Pub. L. No. 107-110, 115 Stat. 1425 (2002) (codified at 20 U.S.C. §§ 6301-7941 (2006)).

<sup>4</sup> See 20 U.S.C. § 6311 (2006).

<sup>5</sup> Improving America’s Schools Act of 1994, Pub. L. 103-382, 108 Stat. 3518 (1994) (codified as amended at 20 U.S.C. § 6301 (2006)).

assesses data from the post-2000 era, arguing that the substantial increases in the rate at which student achievement rose in this era, compared to the previous decade, are due, at least in part, to the flowering of the consequential accountability movement in the late 1990s and with the passage of NCLB in 2001.

In Part IV, the Article discusses several state consequential accountability systems in detail. It seeks to provide the reader with an understanding of some key characteristics of successful accountability systems, as well as insight into how such provisions compare with the provisions of NCLB. Part V highlights some of the commonly cited problems with accountability systems and assesses the validity of these criticisms of consequential accountability. Finally, in Part VI, the Article evaluates the current administration's proposals for the reform of NCLB and offers suggestions for how to maintain the positive aspects of NCLB while modifying aspects that need sharpening.

## II. HISTORY OF EDUCATION REFORM IN THE U.S.

### A. *The Early Evolution of Public Education in America*

The role of public education in American society has changed significantly since its inception in the 17th century. Society's demands on and expectations of the public education system have increased dramatically since the establishment of the first public school in 1635.<sup>6</sup> Under the current system of compulsory public education for all children, the vast majority of Americans take the existence of public education for granted, and the focus on education reform has largely shifted from increasing access to improving quality.

The Boston Latin School, founded in 1635, was the first public school in the United States and is currently the oldest existing public school in the country.<sup>7</sup> Yet, compulsory and free public education did not become mainstream in America until Horace Mann and other leaders of similar mind entered the scene some 200 years later. As the newly appointed Secretary of Education in Massachusetts, Mann observed firsthand the decrepit condition of schools in his state that, ironically, likely had the most progressive public education system in the United States at the time. Horace Mann described the then-existing state of public education in Massachusetts in most unflattering terms:

In this Commonwealth, there are 3,000 public schools, in all of which the rudiments of knowledge are taught. These schools, at

---

<sup>6</sup> *Boston Latin School*, ENCYCLOPEDIA BRITANNICA, available at <http://www.britannica.com/EBchecked/topic/74909/Boston-Latin-School> (last visited Oct. 26, 2010).

<sup>7</sup> *Id.*

the present time, are so many distinct independent communities; each being governed by its own habits, traditions, and local customs. There is no common, superintending power over them; there is no bond of brotherhood or family between them. They are strangers and aliens to each other. As the system is now administered, if any improvement in principles or modes of teaching is discovered by talent or accident, in one school, instead of being published to the world, it dies with the discoverer. No means exist for multiplying new truths, or preserving old ones.<sup>8</sup>

An ever-colorful commentator, Mann is rumored to have suggested that Massachusetts took better care of its livestock than its children.<sup>9</sup> A year after his appointment as the Secretary of Education in Massachusetts, Mann founded and became the editor of *The Common Journal*,<sup>10</sup> which was based on a principle that, among other things, a commonly educated public was necessary—and even “indispensable”—to free society.<sup>11</sup> Mann’s principles eventually triumphed, and now every child in the United States is entitled to—indeed, required to—receive an education.<sup>12</sup>

From 1890 to 1930, the portion of fourteen- to seventeen-year-olds entering high school increased from less than 10% to more than 70%.<sup>13</sup> The development of a common expectation that students across America would enter and hopefully complete high school was a crucial achievement by the middle decades of the twentieth century.

### B. Modern Education Reform

Modern efforts to improve the quality of public education were prompted significantly by two major social and historical forces—the civil rights movement and a growing and widespread concern about the vital importance of education to our national security. This Article picks up the story

<sup>8</sup> LAWRENCE A. CREMIN, *AMERICAN EDUCATION: THE NATIONAL EXPERIENCE, 1783-1876* 155 (1980).

<sup>9</sup> SARAH MONDALE, *SCHOOL: THE STORY OF AMERICAN PUBLIC EDUCATION* 28 (2002).

<sup>10</sup> *Common School Journal* for the year 1839 (Horace Mann ed., 1839).

<sup>11</sup> *See id.* at 5.

<sup>12</sup> States differ on when students are allowed to drop out of school without violating truancy laws. According to an op-ed in the *New York Times* by Barbara Ehrenreich, a growing number of cities have begun ticketing and sometimes handcuffing teenagers found on the streets during school hours. Barbara Ehrenreich, Op-Ed., *Is It Now a Crime to Be Poor?*, *N.Y. TIMES*, Aug. 9, 2009, available at <http://www.nytimes.com/2009/08/09/opinion/09ehrenreich.html?pagewanted=2>. In Los Angeles, the fine for truancy is \$250; in Dallas, it can be as much as \$500. *Id.* There is even at least one website, Hooky Map.com, that displays public data on truancy geographically, highlighting areas that are relatively more or relatively less likely to see a high percentage of truant students. *Florida Hooky Map—School Attendance by Zip—Percent of Florida Students Absent 21 Days +*, HOOKY MAP, <http://www.hookymap.com> (last visited Oct. 26, 2010).

<sup>13</sup> Daniel P. Resnick, *Minimum Competency Testing Historically Considered*, 8 *REV. RES. EDUC.* 3, 17-18 (1980), available at <http://www.jstor.org/stable/1167122?origin=JSTOR-pdf>.

of education reform in the 1950s, when through litigation, and later legislation, the nation became increasingly committed to equal and, ultimately, effective education for children of color.

The Supreme Court's decision in *Brown v. Board of Education*<sup>14</sup>—and the desegregation of schools that followed—has been described as perhaps the most important development in U.S. educational policy of the last century.<sup>15</sup> *Brown* was decided twenty-six years before the establishment of the U.S. Department of Education, which today represents the institutional face of federal involvement in education. However, *Brown* thrust the federal government, largely through federal district courts and the desegregation orders they issued, directly into the policies, budgets, and practices of many school districts across the country.<sup>16</sup> The increasing concern about the education of poor and minority children that grew out of the civil rights period provided much of the impetus needed for the education reform efforts that followed.

In 1965, eleven years after *Brown*, Congress passed the landmark Elementary and Secondary Education Act (“ESEA”), which represents a major piece of civil rights legislation due to its focus on improving the educational opportunities of poor children.<sup>17</sup> This Act, the first of its kind, involved the federal government in the ongoing operations of state and local education affairs, initially through funding districts with schools that primarily serve students from low-income families. Congress originally authorized the ESEA through 1970,<sup>18</sup> but it has been reauthorized repeatedly since.<sup>19</sup>

In the wake of this burst of interest in educating all students, many states began in the 1970s to explore and implement early forms of accountability, largely with a focus on “minimum competency.”<sup>20</sup> That is, states established minimum levels of proficiency for their students, and predicated students' advancement in certain ways on meeting those minimums. For instance, several states required students to pass minimum competency tests in order to advance a grade level or, in some cases, to graduate high school.<sup>21</sup> The aim of this philosophy was to “show the tax-paying public that graduates had been prepared to function at a demonstrable level of competence.”<sup>22</sup>

---

<sup>14</sup> 347 U.S. 483 (1954).

<sup>15</sup> Sarah J. Reber, *School Desegregation and Educational Attainment for Blacks 2* (Nat'l Bureau of Econ. Research, Working Paper No. 13193, 2007), available at <http://www.nber.org/papers/w13193>.

<sup>16</sup> See, e.g., *Cooper v. Aaron*, 358 U.S. 1 (1958) (holding that all school districts were bound by the Supremacy Clause to follow *Brown*, notwithstanding any state laws or policies to the contrary; the case stemmed from the Little Rock Nine incident); *Griffin v. County School Board of Prince Edward County*, 377 U.S. 218 (1964) (requiring a county to levy taxes to keep its public schools open in the name of desegregation).

<sup>17</sup> Elementary and Secondary Education Act, Pub. L. No. 89-10, 79 Stat. 27 (1965) (codified as amended at 20 U.S.C. §§ 6301-7941 (2006)).

<sup>18</sup> 79 Stat. at 36.

<sup>19</sup> See, e.g., No Child Left Behind Act of 2001, Pub. L. No. 107-110, 115 Stat. 1425 (2002) (codified at 20 U.S.C. §§ 6301-7941 (2006)) (latest reauthorization).

<sup>20</sup> *Id.* at 4.

<sup>21</sup> See Resnick, *supra* note 13, at 8-9.

<sup>22</sup> *Id.* at 9.

As a contemporaneous journal article described it, “minimum competency tests are intended to return meaning to the high school diploma by requiring that students meet various indicia of basic competence.”<sup>23</sup>

The publication of *A Nation at Risk* (“*Nation*”) by the U.S. Department of Education in 1983 brought increased attention to education reform in the United States.<sup>24</sup> This landmark report described in the direst of terms the failings of the educational system, and, playing upon Cold War-era fears, analogized its potential detrimental effects to that of a foreign act of war.<sup>25</sup> It argued that the nation’s prosperity was imperiled and implied that other nations with better-educated populaces would overtake the U.S. economy if the education system were not reformed.<sup>26</sup>

*Nation* characterized the need to raise the level of education in the United States as a matter of national security, and called for a shift away from “minimum competency” toward a more “input-focused” education reform approach that would lead to a much better educated citizenry.<sup>27</sup> Whereas “minimum competency” focused primarily on showing minimal levels of student proficiency, “input-focused” reform was geared toward setting high expectations and assuring schools certain resources, such as more money and more teachers, to help students attain these standards of proficiency. The impact of the report was enormous, both in terms of its effect on education reform philosophy and its impact on the role of the federal government in education policy.<sup>28</sup> During its first year, 70,000 copies of *A Nation at Risk* were sold, and approximately seven times that number were produced and distributed.<sup>29</sup>

States and districts enacting *Nation*-influenced reforms increased school budgets, decreased student/teacher ratios, increased credit requirements for graduation, increased the number of science classes students were required to take, lengthened schools days and school years, and raised

<sup>23</sup> Lorrie Shepard, *Technical Issues in Minimum Competency Testing*, 8 REV. RES. EDUC. 30, 30 (1980), available at <http://www.jstor.org/stable/1167123?origin=JSTOR-pdf>.

<sup>24</sup> DIANE MASSELL & SUSAN FUHRMAN, TEN YEARS OF STATE EDUCATION REFORM, 1983-1993: OVERVIEW WITH FOUR CASE STUDIES v (1994).

<sup>25</sup> See NAT’L COMM’N ON EXCELLENCE IN EDUC., A NATION AT RISK: THE IMPERATIVE FOR EDUCATIONAL REFORM 5 (1983) [hereinafter *NATION AT RISK*], available at <http://reagan.procon.org/sourcefiles/a-nation-at-risk-reagan-april-1983.pdf> (“If an unfriendly foreign power had attempted to impose on America the mediocre educational performance that exists today, we might well have viewed it as an act of war. As it stands, we have allowed this to happen to ourselves. . . . We have, in effect, been committing an act of unthinking, unilateral educational disarmament.”).

<sup>26</sup> Diane Ravitch, *A Historic Document*, in OUR SCHOOLS & OUR FUTURE: ARE WE STILL AT RISK? 28 (Paul E. Peterson ed., 2003).

<sup>27</sup> See MASSELL & FUHRMAN, *supra* note 24, at 5.

<sup>28</sup> *Id.* at 2 (noting that, despite huge state budget deficits, state legislatures “eagerly” enacted *Nation*-inspired reforms; for example, between 1983 and 1985, 41 states increased high school graduation requirements, a “diffusion rate” more than four times the historical rate for state policies without specific federal impetus).

<sup>29</sup> *Id.* at 1.

teacher salaries, among other measures.<sup>30</sup> The report's impact was such that an extraordinary number of states authorized commissions to examine and propose implementation of many of its policy recommendations, notwithstanding budget constraints due to a recession.<sup>31</sup>

The impact that the 1983 *Nation* report had on the federal government's role in education is somewhat ironic. The Reagan administration, which took office two years prior to the release of *Nation*, had intended to dismantle the U.S. Department of Education.<sup>32</sup> After *Nation*'s release, however, it was politically unfeasible to dismantle the Department and, although *Nation* was not accompanied by large new federal programs,<sup>33</sup> the report "end[ed] the debate about abolishing the department and guaranteed its political survival."<sup>34</sup> Having survived the Reagan administration, the federal government's role in education policy grew with each successive administration.

Beginning in the late 1980s, and with increasing momentum in the 1990s, another dramatic shift occurred in education policy as a result of disillusionment with the sparse results of "input-focused" reform.<sup>35</sup> This seismic philosophical shift changed the focus from reliance on certain inputs to a more results-oriented approach, in which the focus was on the achievement of improved student results, or what may be called outputs.

During this era, a new direction in policy emerged, one based on standards-based reform, that is, a systemic approach to school policy in which critical attention is paid to: (a) the standards by which students are taught, (b) the measurement and disclosure of how well students learn to those standards, and (c) in some states, accountability for success or failure in achieving improved learning. According to Eric Hanushek and Margaret Raymond, these three facets of standards-based reform are the backbone of "consequential accountability," the focus of this Article.<sup>36</sup>

The standards-based reform movement began at the state level. By the 1994-1995 school year, virtually every state had developed explicit standards of achievement and had implemented tests that were aligned with those standards.<sup>37</sup> Some states focused mostly or exclusively on implementing the first two elements, putting emphasis on the articulation of curriculum standards and testing against those standards, principally for diagnostic purposes. In these states, testing was only loosely connected to accountability; it was used mainly to divide students into academic tracks or for diagnostic

---

<sup>30</sup> NATION AT RISK, *supra* note 25, at 13-20.

<sup>31</sup> See MASSELL & FUHRMAN, *supra* note 24, at 2.

<sup>32</sup> Ravitch, *supra* note 26, at 27.

<sup>33</sup> MASSELL & FUHRMAN, *supra* note 24, at 1.

<sup>34</sup> Terrell H. Bell, *Reflections One Decade After "A Nation at Risk"*, 74 PHI DELTA KAP-  
PAN 592, 592-97 (1993); Ravitch, *supra* note 26, at 27.

<sup>35</sup> See MASSELL & FUHRMAN, *supra* note 24, at 3-5 (describing how the *Nation* reforms led to an increase in inputs but not necessarily an increase in results).

<sup>36</sup> Hanushek & Raymond, *supra* note 2, at 306.

<sup>37</sup> DIANE MASSELL ET AL., PERSISTENCE AND CHANGE: STANDARDS-BASED REFORM IN NINE STATES 3 (1997).

purposes.<sup>38</sup> One might refer to these states as “pure standards and testing” states.

Other states, however, used tests for more than just diagnostic or tracking purposes. In addition to having standards and assessments, these states made a fundamental change in policy and attached *consequences* for schools based on the assessment results.<sup>39</sup> These “consequential accountability” states used tests to sanction schools that performed poorly and to reward schools that performed well.

The rationale behind such a system is as follows: by attaching incentives to increases in student achievement, schools will be more motivated to effectuate such improvements. The novelty of the approach was that, for the first time, there would be consequences for *schools* for their performance. Test results have sometimes had consequences for *students*, and such testing was not new. For example, New York has used Regents examinations to test students’ command of high school curriculum since the 19th century,<sup>40</sup> and the Iowa Test of Basic Skills (“ITBS”) dates back to 1935.<sup>41</sup> However, these tests were primarily diagnostic: “[h]ow well students, classes, or schools performed on [them] had few consequences [for the school itself].”<sup>42</sup>

Texas was among the pioneers of the consequential accountability movement, implementing a statewide accountability system in 1993.<sup>43</sup> Under its system, Texas used a state assessment to measure school performance directly so that it could sanction those schools not meeting performance targets and reward schools exceeding targets.<sup>44</sup> By 1996, twelve states employed some version of a “consequential accountability” model and by 2000 that number had grown to thirty-nine.<sup>45</sup> The states that began adopting the philosophy that schools should be held accountable for the progress (or lack thereof) of their students laid much of the groundwork for the eventual passage of NCLB, the current high watermark of federal involvement in education policy.

The federal government also began to move toward consequential accountability in the 1990s. George H.W. Bush was elected in 1988 promising to be “the education president.”<sup>46</sup> Commensurate with his promise, the trend toward decreasing the federal government’s share of education expenditure

---

<sup>38</sup> Martin Carnoy & Susanna Loeb, *Does External Accountability Affect Student Outcomes? A Cross-State Analysis*, 24 *EDUC. EVALUATION & POL’Y ANALYSIS* 305, 306 (2002).

<sup>39</sup> Hanushek & Raymond, *supra* note 2, at 1-2.

<sup>40</sup> Carnoy & Loeb, *supra* note 38, at 306.

<sup>41</sup> *Id.*

<sup>42</sup> *Id.*

<sup>43</sup> *Id.*

<sup>44</sup> *Id.*

<sup>45</sup> Hanushek & Raymond, *supra* note 2, at 298. States “diverge[d] considerably” in the types of consequences attached to performance. *Id.* at 310. This Article discusses the different consequential accountability models in depth. *See infra* Part IV.

<sup>46</sup> PBS, *Domestic Politics*, AMERICAN EXPERIENCE GEORGE H.W. BUSH, <http://www.pbs.org/wgbh/amex/bush41/more/domestic.html> (last visited Oct. 26, 2010).



reversed and, in fact, its share increased, albeit slightly, to 6.3%.<sup>47</sup> Additionally, the first year of the President's term included a historic "Education Summit" that led to a set of new, bipartisan national educational goals, the first of its kind since *Nation*.<sup>48</sup>

In 1994, Congress passed and President Clinton signed into law the Improving America's Schools Act ("IASA"). For the most part, IASA merely consolidated several existing federal education programs and created new grants for state programs focused on areas such as professional development and English proficiency.<sup>49</sup> However, IASA introduced an early version of what would be a defining provision of NCLB: federally mandated use of state assessments to measure students' progress toward state standards.<sup>50</sup> These state assessments were to be used to identify those schools in need of improvement. Schools that continuously performed poorly were identified for "corrective action."<sup>51</sup> In practice, though, this mechanism did little to remedy the ills responsible for these schools' poor performance. Other than placement on a list of schools identified as in need of "corrective action," few consequences were attached to poor performance; worse still, some schools did not even know they were on such lists at all.<sup>52</sup>

NCLB, signed into law by President George W. Bush after less than a year in office,<sup>53</sup> sought to remedy the failings of IASA's early effort at identifying problem schools. NCLB strengthened the accountability provisions of IASA while maintaining the basic premise that states should come up with the standards themselves and provide assessments against those standards. In sum, NCLB took a consequential accountability model that theretofore had existed in approximately 39 states and sought to implement it nationally.<sup>54</sup> At the time that it was passed, the No Child Left Behind Act of 2001 represented the high point of the federal government's involvement in efforts to assure and improve the quality of American education.

NCLB required schools in every state to bring their students up to full "proficiency"<sup>55</sup> in math and reading by 2014, but it allowed each state to devise its own plan as to how to get there.<sup>56</sup> Some states' plans called for

---

<sup>47</sup> MASSELL & FUHRMAN, *supra* note 24, at 11.

<sup>48</sup> *Id.* at 19-20.

<sup>49</sup> See generally U.S. DEP'T OF EDUC., THE IMPROVING AMERICA'S SCHOOLS ACT OF 1994: REAUTHORIZATION OF THE ELEMENTARY AND SECONDARY EDUCATION ACT (1995), available at <http://eric.ed.gov/ERICWebPortal/contentdelivery/servlet/ERICServlet?accno=ED399649>.

<sup>50</sup> *Id.*

<sup>51</sup> Sara Mead, *Easy Way Out: "Restructured" Usually Means Little Has Changed*, 7 EDUC. NEXT 53 (Winter 2007), available at [http://educationnext.org/files/ednext\\_20071\\_52.pdf](http://educationnext.org/files/ednext_20071_52.pdf).

<sup>52</sup> *Id.*

<sup>53</sup> President George W. Bush, Statement Upon Signing H.R. 138 (Jan. 8, 2002), in 2002 U.S.C.C.A.N. 1614.

<sup>54</sup> See Hanushek & Raymond, *supra* note 2, at 298.

<sup>55</sup> Specifically, NCLB required that 95% of all students in a school, as well as 95% of each principal subgroup of students, be proficient in math and reading by 2014. 20 U.S.C. §§ 6311(b)(2)(C)(v)(II), 6311(b)(2)(I)(ii) (2006) (95% testing requirement).

<sup>56</sup> See 20 U.S.C. § 6311 (2006) (discussing use of "State plans" and the requirements thereof, but containing much open-ended language allowing states flexibility).

fairly consistent growth in test scores while other states' plans called for slow initial growth with rapid growth closer to 2014.<sup>57</sup> Under NCLB, schools that fail to meet their goal for a given year are said to have not made Adequate Yearly Progress ("AYP").<sup>58</sup> Schools that do not make AYP are identified as needing improvement and are subject to a series of escalating interventions.<sup>59</sup> For these low-performing schools, the interventions begin with school choice and supplemental tutoring for students.<sup>60</sup> If a school continuously fails to make AYP for five years, it must develop a restructuring plan, which goes into action if the school fails to make AYP for a sixth year. The restructuring options include: (1) closing and reopening as a charter; (2) replacing relevant staff; (3) turning governance over to the state; (4) contracting with a private management company to operate the school; and (5) "any other major restructuring of the school's governance designed to produce major reform."<sup>61</sup>

In many respects, the Obama administration extended these policies, pursuing an active education agenda in its first two years. Control of Congress, however, has recently shifted substantially in the mid-term elections, seemingly away from policies that favor a significant federal role. Reauthorization of the ESEA is also now four years overdue. In this uncertain environment, efforts to modify federal policy would benefit from an analysis of the general effectiveness of accountability systems on student achievement, including a close examination of legislation and policy at both the state and federal levels. The next Part of this Article will provide just that: an analysis of (1) the effectiveness of state regimes that enacted consequential accountability systems in the 1990s vis-à-vis states that did not and (2) the positive effects on student achievement of federal consequential accountability efforts implemented in the late 1990s and culminating with the passage of NCLB in 2001.

### III. THE EFFECTIVENESS OF CONSEQUENTIAL ACCOUNTABILITY

#### A. *State Accountability Systems in the 1990s*

Historically, comparisons of student achievement across states have been difficult to make. There are multiple differences in important charac-

---

<sup>57</sup> See generally *Approved State Accountability Plans*, U.S. DEP'T OF EDUC., <http://www2.ed.gov/admins/lead/account/stateplans03/index.html> (last updated Aug. 2010).

<sup>58</sup> 20 U.S.C. § 6311 (2006); 34 C.F.R. § 200.13-200.14 (2007).

<sup>59</sup> 20 U.S.C. § 6316(b) (2006).

<sup>60</sup> Mead, *supra* note 51, at 52.

<sup>61</sup> 20 U.S.C. § 6316(b)(8)(B) (2006); Mead, *supra* note 51, at 52. This "fifth option" is a "safety valve" provision for schools that have heretofore often foreclosed the more stringent accountability options. See *infra* Part IV.B for a more elaborate discussion.

teristics of education in the states, for example: crucial family variables;<sup>62</sup> pupil-teacher ratios;<sup>63</sup> provision of pre-K programs;<sup>64</sup> average teacher experience;<sup>65</sup> average salary levels;<sup>66</sup> and per-pupil expenditures.<sup>67</sup> Lack of uniform measures of success compounds this difficulty. While tests like the SAT and ACT are given nationally, vastly different participation rates render them unreliable as comparative assessments,<sup>68</sup> and the fact that the tests are typically only taken by eleventh and sometimes twelfth graders precludes comparing year-to-test results.

At least the Department of Education alleviated the uniform measures problem when it began giving the National Assessment of Educational Progress (“NAEP”) test to representative samples of students across the states.<sup>69</sup> NAEP test results have since become integral to the vast majority of educational reform studies. While comparative state data did not exist prior to the 1990s until the implementation of NAEP exams that provided state-level data, a plethora of *national* data for the period from the 1970s through the 1990s, as well as non-uniform state data, exists, allowing researchers to evaluate the accountability movement in all its phases and draw conclusions about its impact.

Eric Hanushek and Margaret Raymond’s statistical analysis suggests that accountability had a positive impact on student performance in states that adopted consequential accountability systems in the 1990s. Specifically, states with consequential accountability tended to show more rapid gains in NAEP scores than states without.<sup>70</sup> Martin Carnoy and Susanna Loeb’s research corroborates Hanushek and Raymond’s findings by demonstrating that students in high-accountability states—states with very *strong* accountability systems that attach more stringent consequences to results—averaged significantly greater gains on the NAEP eighth grade math test than students in states without strong accountability programs.<sup>71</sup> Their research shows that states that implemented accountability systems in the 1990s saw larger gains in NAEP performance between 1996 and 2000 than states that did not have accountability systems.<sup>72</sup> The results were robust to numerous specification

---

<sup>62</sup> DAVID W. GRISSMER ET AL., IMPROVING STUDENT ACHIEVEMENT: WHAT STATE NAEP TEST SCORES TELL US xx1, xxiii, 97 (2000), available at [http://www.rand.org/pubs/monograph\\_reports/MR924/index.html](http://www.rand.org/pubs/monograph_reports/MR924/index.html).

<sup>63</sup> *Id.*

<sup>64</sup> *Id.* at xxvi.

<sup>65</sup> *Id.* at xviii.

<sup>66</sup> *Id.*

<sup>67</sup> *Id.*

<sup>68</sup> *Id.* at 3.

<sup>69</sup> *Id.* at 4; see also Hanushek & Raymond, *supra* note 2, at 304 (using NAEP data for further analysis). NAEP, prior to 1990, had been given to a representative sample of the *country*, known as “NAEP long-term trends,” but it was only in this year that comparing states became possible.

<sup>70</sup> Hanushek & Raymond, *supra* note 2, at 309-10.

<sup>71</sup> Carnoy & Loeb, *supra* note 38, at 309.

<sup>72</sup> *Id.* at 322.

checks, including controls for exclusion of students from tests due to classification as in need of special education or having limited English proficiency.<sup>73</sup> In particular, there was a positive and significant relationship between the strength of states' accountability system and math achievement gains at the eighth grade level across racial/ethnic groups.<sup>74</sup>

In another study, the RAND Corporation similarly linked consequential accountability to increases in student achievement, and specifically found that the two states that pioneered the earliest consequential accountability models, North Carolina and Texas, had the largest NAEP gains during that time period.<sup>75</sup> Moreover, according to the RAND paper, several factors suggest that accountability reforms were the greatest cause of NAEP gains across the states in the 1990s.<sup>76</sup> First, major changes in resources cannot explain these gains.<sup>77</sup> Also, a case study of Texas and North Carolina suggests that the most plausible cause of the gains in these states were their reform policies, which aligned standards, assessments, and accountability.<sup>78</sup> Other things being equal, higher per-pupil expenditures, lower pupil-teacher ratio in lower grades, higher reported adequacy of teacher-reported resources, higher levels of participation in public pre-kindergarten, and lower teacher turnover all showed positive, statistically significant effects on achievement.<sup>79</sup> However, according to RAND, these variables explain one-half or less of the non-family difference in achievement across states.<sup>80</sup> While family factors explain the largest discrepancies in achievement, the RAND report suggests that most of the "non-family" differences are attributable to consequential accountability reforms.<sup>81</sup>

The most convincing evidence, however, in support of consequential accountability is that states that only embrace standards or assessments without accountability do not show the gains that full-fledged accountability states do. As stated earlier, Hanushek and Raymond define consequential accountability systems as being comprised of three elements: (1) standards; (2) assessments; and (3) accountability.<sup>82</sup> States that embraced only the first one or two elements *did not* show statistically significant gains in student

<sup>73</sup> *Id.*

<sup>74</sup> *Id.* at 320.

<sup>75</sup> GRISSMER ET AL., *supra* note 62, at 99-100.

<sup>76</sup> *Id.*

<sup>77</sup> *Id.* That is, according to the Grissmer report, there was no statistically significant correlation sufficient to suggest relationship between NAEP success and resource-input differentials. *Id.*

<sup>78</sup> *Id.*

<sup>79</sup> *Id.* Other variables deemed important to improving student achievement under conventional wisdom—higher teacher salaries, higher teacher educational levels, and increased experience—did not show significant effects on achievement. *Id.*

<sup>80</sup> *Id.* at 98.

<sup>81</sup> *Id.* at 99-100. That said, the RAND report also indicates that nothing that government has done has shown greater effects than "family factor" differentials, which just goes to the limitations on government's ability to effectuate educational achievement on its own.

<sup>82</sup> *See generally id.*

achievement, whereas states that embraced all three elements did show such gains.<sup>83</sup>

As noted above, every state had “standards.” Although there was variation in their implementation, states with standards that were implemented rapidly and aggressively did not perform better in any statistically significant way than did those with more incrementally-implemented standards,<sup>84</sup> which suggests that the standards *alone* mean little. Moreover, states that had assessments in addition to standards (but no accountability) also did not do particularly well. For example, states with “report card” systems—that is, where parents get a report card describing how the school is performing based on test results compared against some set of standards—did not perform better statistically than states without.<sup>85</sup>

In sharp contrast, states with standards, assessments, *and* accountability<sup>86</sup> have performed significantly better than states that are lacking any one of these elements.<sup>87</sup> In fact, outside of family factors,<sup>88</sup> the presence or lack of an accountability system appears to have had the most significant effect on student achievement in the 1990s.<sup>89</sup>

### B. Consequential Accountability Post-2000

Since the number of states utilizing consequential accountability had grown from three in 1993 to thirty-nine by 2000,<sup>90</sup> this Article will compare student achievement from the decade prior to the pervasive spread of these policies to the decade after they were broadly in place. This analysis supports the cited research in demonstrating the impact that the consequential accountability policies of the late 1990s and early 2000s had on student achievement. This Article will focus on 1999 as a turning point because, by that year, a substantial majority of states had adopted significant consequential accountability policies.<sup>91</sup> Consequently, the years following 1999 represent the time period in which consequential accountability policies were firmly ingrained across the nation, a trend that continued with the passage of NCLB in 2001. By contrast, in the early and mid-1990s, less than a handful of states had even begun to experiment with consequential accountability.

---

<sup>83</sup> *Id.*

<sup>84</sup> MASSELL ET AL., *supra* note 37, at 25.

<sup>85</sup> Hanushek & Raymond, *supra* note 2, at 310.

<sup>86</sup> That is, those with “consequential accountability” systems.

<sup>87</sup> See Hanushek & Raymond, *supra* note 2, at 321.

<sup>88</sup> Because family factors are difficult to ascertain, and therefore to control for, this Article does not address them. However, it is worth noting that family factors, such as whether both parents are in the home or the education level of the parents, have a stronger effect on student achievement than anything government has done. See GRISSMER ET AL., *supra* note 62, at 97-98. This highlights the limitations on the government’s ability to solve the problems facing the United States education system.

<sup>89</sup> *Id.* at 98.

<sup>90</sup> Hanushek & Raymond, *supra* note 2, at 298.

<sup>91</sup> *Id.*

Much of the analysis of post-2000 data performed by others is done in the context of assessing the impact of the No Child Left Behind Act of 2001 on student achievement.<sup>92</sup> While these types of studies provide a helpful context for assessing the effectiveness of consequential accountability, the results of such studies are both mixed and fairly inconclusive.

Because NCLB went into effect so recently, researchers lack long-term data, an important element in analyzing whether NCLB has had a positive impact on student achievement. Another dilemma that researchers face is in determining which data points should be used to assess the effectiveness of NCLB. Many researchers have noted that the nation experienced significant gains in student achievement in 2002 and 2003,<sup>93</sup> which requires a determination of whether, and to what extent, these gains should or should not be attributed to NCLB.

In contrast, it is much more helpful to examine the effects of consequential accountability generally (rather than NCLB, specifically) post-2000, since this was the period in which one would expect to see the impact of the increased use of consequential accountability by states in the 1990s. Because consequential accountability was gaining momentum in the late 1990s, and reached its height with the passage of NCLB in 2001, one can expect that the growth in student achievement in the early 2000s, which continued throughout the decade, may have been caused, at least in part, by the rise of consequential accountability policies across the nation. Put another way, it is less helpful to consider whether the growth in student achievement in the early 2000s is directly attributable to NCLB than it is to consider the effects of the growth of consequential accountability in the 1990s and its expansion in the 2000s, extending through NCLB.

Therefore, this Article will focus more broadly on the effects of consequential accountability in general. NCLB is but one chapter of the story of consequential accountability, as evidenced by the fact that by the late 1990s nearly forty states had implemented consequential accountability systems. Thus, after reviewing studies of NCLB, this Article will argue that when one examines current data from the NAEP Long-Term Trend Assessment, it appears that consequential accountability had a substantial positive impact on student achievement in the 2000s.<sup>94</sup>

---

<sup>92</sup> See, e.g., JAEKYUNG LEE, TRACKING ACHIEVEMENT GAPS AND ASSESSING THE IMPACT OF NCLB ON THE GAPS: AN IN-DEPTH LOOK INTO NATIONAL AND STATE READING AND MATH OUTCOME TRENDS 10-11 (2006), available at <http://civilrightsproject.ucla.edu/research/k-12-education/integration-and-diversity/tracking-achievement-gaps-and-assessing-the-impact-of-nclb-on-the-gaps/lee-tracking-achievement-gaps-2006.pdf>; Bruce Fuller et al., *Gauging Growth: How to Judge No Child Left Behind?*, 36 EDUC. RESEARCHER 268, 275 (2007), available at [http://www.aera.net/uploadedFiles/Publications/Journals/Educational\\_Researcher/3605/07EDR07\\_268-278.pdf](http://www.aera.net/uploadedFiles/Publications/Journals/Educational_Researcher/3605/07EDR07_268-278.pdf).

<sup>93</sup> See, e.g., LEE, *supra* note 92.

<sup>94</sup> This Article is not intended to provide a rigorous scientific analysis of the data. Thus, while the data appear to support the hypothesis that consequential accountability had a positive effect on student achievement in the later 1990s and early 2000s, more research is needed to evaluate this conclusion and to control for other factors that may be at work. While the authors

### 1. Brief Review of Prior Studies

In one of the most recent statistical studies of NCLB, Dee and Jacob found that NCLB generated statistically significant improvements in fourth grade math scores on the main NAEP assessment, as well as some improvement in eighth grade math scores, particularly among low-performers.<sup>95</sup> However, they found no evidence that NCLB increased reading achievement in either fourth or eighth grade.<sup>96</sup> In conducting their analysis, Dee and Jacob compared trends in student achievement for groups of states that had consequential accountability systems similar to that of NCLB prior to the implementation of NCLB with other groups of states that either did not have accountability policies prior to the enactment of NCLB or that had accountability policies that did not closely resemble NCLB prior to its enactment.<sup>97</sup> Dee and Jacob found that NCLB did have some positive effects on student achievement.<sup>98</sup> Importantly, they note that their study may actually have *underestimated* the effects of accountability policies on student achievement.<sup>99</sup>

In contrast, several researchers have examined data from the main NAEP assessment and have concluded that NCLB had little or no effect on student achievement.<sup>100</sup> These studies have struggled with the aforementioned issue of whether the gains experienced in 2002 and 2003 should be considered pre-NCLB gains or post-NCLB gains. Again, this Article will not attempt to resolve this ongoing debate. Rather, it will attempt to show that when one examines current NAEP Long-Term Trend data,<sup>101</sup> substantial evidence supports the notion that consequential accountability's rise in the late

---

do not attempt to demonstrate scientifically the causal impact of consequential accountability policies on student achievement, they believe that the correlation between the rise of consequential accountability and the gains in student achievement during the years that followed supports the hypothesis that consequential accountability leads to positive gains in student achievement.

<sup>95</sup> Thomas Dee & Simon Jacob, *The Impact of No Child Left Behind on Student Achievement* 36-37 (Nat'l Bureau of Econ. Research, Working Paper No. 15531, 2009), available at <http://www.nber.org/papers/w15531>.

<sup>96</sup> *Id.* at 37.

<sup>97</sup> *Id.* at 16.

<sup>98</sup> *Id.* at 20.

<sup>99</sup> *Id.* The reason Dee and Jacob suggest that they may have underestimated gains under NCLB is that their estimates identified the impact of NCLB-induced school-accountability provisions on states *without* prior accountability policies, thus leaving out states that adopted accountability policies prior to NCLB. *Id.*

<sup>100</sup> For example, in a 2006 comprehensive study, Lee concludes that NCLB had little effect on student achievement and did not result in a significant increase in the rate at which achievement gaps are narrowing. LEE, *supra* note 92, at 10-11. In a 2007 study, Fuller, Wright, Gesicki, and Kang describe NCLB as having mixed results. The researchers note that on the NAEP assessment, significant progress occurred in both reading and math around the time that NCLB went into effect, but contend that these gains cannot be attributed to NCLB since they occurred before NCLB had any real chance to impact student achievement. Fuller et al., *supra* note 92, at 275.

<sup>101</sup> The NAEP data is examined in the context of the 1990s growth of consequential accountability rather than in the context of NCLB's passage.

1990s and the 2000s is correlated with gains in student achievement and is a likely cause of these gains.

## 2. Methodology

This Article's examination of the impact of consequential accountability on student achievement in the current decade primarily uses data from the NAEP Long-Term Trend Assessment, an assessment that was designed to allow for a valid comparison of data over the entire period for which the NAEP has been in existence.<sup>102</sup> The discussion of scores of English language learners and students with disabilities is based on data from the main NAEP assessment, since the Long-Term Trend Assessment does not disaggregate data for these groups.<sup>103</sup>

In order to address the possibility that student test scores were rising at similar rates both before and after consequential accountability became widespread, this Article will compare score trends from 1990-1999 with those from 1999-2008, examining differences in the *rates* at which scores changed before and after consequential accountability became widespread.<sup>104</sup>

## 3. Effect on Reading Scores

### a. General

Reading scores on the NAEP Long-Term Trend Assessment have increased since 2004 for all three of the age groups included in the test: 9-, 13-, and 17-year-old students. On the NAEP exam, a gain of ten points corresponds to roughly one grade level.<sup>105</sup> From 1990 to 1999, 9-year-old reading scores increased from 209 to 212 points, a gain of only 3 points; in contrast, from 1999 to 2008, 9-year-old reading scores increased by 8

---

<sup>102</sup> Fuller et al., *supra* note 92, at 269.

<sup>103</sup> The key distinction between the NAEP's disaggregated data and the NCLB's disaggregated data is that the NAEP exam is only given to a relatively small sample of students within each state and only provides state-level and national-level data, whereas state tests under NCLB provide school-level data. Moreover, NCLB's data disaggregation requirement is significant because it required states to collect disaggregated data at the school level for the first time and then to hold schools responsible for the performance of all of their subgroups of students.

<sup>104</sup> See Neal McCluskey & Andrew J. Coulson, *End It, Don't Mend It: What to Do with No Child Left Behind*, CATO INST. POL'Y ANALYSIS, Sept. 5, 2007, at 1, 3, available at <http://www.cato.org/pubs/pas/Pa599.pdf> (noting that pre-NCLB score trends must be compared with post-NCLB score trends in order to "have any hope of isolating NCLB's actual effect on student achievement and test score gaps").

<sup>105</sup> U.S. DEPT. OF EDUC., SCHOOL POVERTY AND ACADEMIC PERFORMANCE: NAEP ACHIEVEMENT IN HIGH-POVERTY SCHOOLS—A SPECIAL EVALUATION REPORT FOR THE NATIONAL ASSESSMENT OF TITLE I (Sept. 1998), <http://www2.ed.gov/pubs/schoolpoverty/index.html>.



(11\*)<sup>106</sup> points, rising from 212 to 220 points.<sup>107</sup> Reading scores for 13- and 17-year-olds remained fairly stagnant during this period.<sup>108</sup>

FIGURE 1. CHANGES IN NAEP LONG-TERM TREND READING SCORES.

	All Scores		Black Scores		Hispanic Scores	
	1990– 1999	1999– 2008	1990– 1999	1999– 2008	1990– 1999	1999– 2008
9-year-old	+3	+8	+4	+21	+4	+14
13-year-old	+2	+1	-3	+9	+6	-2
17-year-old	-2	-2	-3	+2	-4	-2

*b. Gains by Race*

As shown in Figure 1, when reading scores on the NAEP Long-Term Trend Assessment are broken down by race, the results are impressive. Scores for black 9-year-olds increased from 186 to 204 from 1999 to 2008, an astounding 18 (21\*) point increase.<sup>109</sup> By comparison, from 1990 to 1999, reading scores for black 9-year-olds increased by only 4 points, from 182 to 186.<sup>110</sup> Even more impressively, the black-white achievement gap in reading among 9-year-olds closed from 35 points in 1999 to 24 points in 2008, even as all white scores went up.<sup>111</sup> During the previous nine years, in contrast, this black-white disparity had shown no change.<sup>112</sup> The incredible growth witnessed in the decade in which consequential accountability became widespread nearly equaled that of the 1970s, when the struggles of the civil rights

<sup>106</sup> The number in parentheses (“11\*”) represents the change in reading scores if one accounts for the changes in the NAEP test that occurred in 2004. The NAEP Long Term Trend Assessment underwent several changes in 2004 and data are available in 2004 for both the old and new versions of the test. This Article uses two methods when calculating changes in scores from 1999 to 2008. The primary method is to simply subtract the 1999 score from the 2008 score, a method that may not fully capture the difference in scores since it does not account for variations in the 2004 scores between the two versions of the test. The second method calculates the change in scores for each version of the test and adds them together (i.e.  $(2004_{old} - 1999_{old}) + (2008_{new} - 2004_{new})$ ). This result is indicated by including the second calculation in parentheses with an asterisk.

<sup>107</sup> BOBBY D. RAMPEY ET AL., NAT’L CTR. FOR EDUC. STATISTICS, U.S. DEP’T OF EDUC. INST. OF EDUC. SCIS., THE NATION’S REPORT CARD: 2008 TRENDS IN ACADEMIC PROGRESS: EXECUTIVE SUMMARY 3 (2009), available at <http://nces.ed.gov/nationsreportcard/pubs/main2008/2009479.asp>.

<sup>108</sup> *Id.* at 3. Reading scores for 13-year-olds increased from 257 to 259 from 1990 to 1999; from 1999 to 2008, the scores increased from 259 to 260 points, a gain of only 1 (3\*) point. Reading scores for 17-year-olds decreased by 2 points from 1990 to 1999 and decreased by 2 (0\*) points from 1999 to 2008. *Id.*

<sup>109</sup> *Id.* at 14.

<sup>110</sup> *Id.*

<sup>111</sup> *Id.*

<sup>112</sup> *Id.* The white-black reading achievement gap for 9-year-olds was 35 points in both 1990 and 1999.

movement began yielding greater equality in educational opportunity among the races.<sup>113</sup>

From 1988 to 1996, well before all states were required to implement consequential accountability systems, the reading scores of black 13-year-olds actually declined, dropping from 243 to 234 points.<sup>114</sup> Yet, between 1999 and 2008, scores for this group rose from 238 to their highest point ever thus far, 247.<sup>115</sup> This 9 (4\*) point increase during the 2000s marks an incredible achievement in light of the fact that from 1990 to 1999, reading scores for black 13-year-olds continued the downward trajectory that had begun in 1988, decreasing by a total of 3 points during this period.<sup>116</sup> Additionally, the black-white achievement gap in reading for 13-year-olds closed by 8 points from 1999 to 2008, reversing an 8-point increase between 1990 and 1999.<sup>117</sup>

Black 17-year-olds only saw a 2 (4\*) point rise in their reading scores from 1999 to 2008.<sup>118</sup> Because white scores for 17-year-olds remained the same during this period, the black-white achievement gap closed by 2 points.<sup>119</sup> From 1990 to 1999, however, black and white scores declined by 3 and 2 points, respectively.<sup>120</sup> Thus, while there is still work to be done in our high schools, the NAEP scores do generally show signs of progress during the 2000s.

Scores for Hispanic 9-year-olds remained fairly stagnant from 1975 to 1999, staying in the range of 183 to 194.<sup>121</sup> In 2008, however, their scores reached their all-time high of 207.<sup>122</sup> This rise in scores reflects a gain of 14 (20\*) points from 1999 to 2008.<sup>123</sup> By comparison, from 1990 to 1999, reading scores for Hispanic 9-year-olds only increased by 4 points.<sup>124</sup> Furthermore, while the Hispanic-white achievement gap in reading for 9-year-olds did not narrow between 1990 and 1999, this achievement gap closed by 7 points between 1999 and 2008, while scores for both whites and Hispanics rose.<sup>125</sup> On the negative side, reading scores for Hispanic 13- and 17-year-olds decreased from 1999 to 2008.<sup>126</sup>

---

<sup>113</sup> *Id.* From 1971 to 1980, 9-year-old black students' scores increased by 19 points on the NAEP long term reading assessment and the racial achievement gap narrowed by 12 points.

<sup>114</sup> *Id.* at 15.

<sup>115</sup> *Id.*

<sup>116</sup> *Id.*

<sup>117</sup> *Id.*

<sup>118</sup> *Id.*

<sup>119</sup> *Id.*

<sup>120</sup> *Id.*

<sup>121</sup> *Id.* at 16.

<sup>122</sup> *Id.*

<sup>123</sup> *Id.*

<sup>124</sup> *Id.*

<sup>125</sup> *Id.*

<sup>126</sup> *Id.* at 17.

c. *English Language Learners*

English language learners (“ELLs”) demonstrated some signs of improvement in reading during the 2000s. There is a lack of long-term data for the NAEP scores of ELLs because data are only available beginning in 1998; consequently, it is difficult to assess the significance of the improvements in ELL scores since 2000.<sup>127</sup> Nonetheless, the available data show signs that consequential accountability likely had a positive impact on the performance of ELLs on the NAEP Assessment.<sup>128</sup>

In fourth grade reading, for example, ELL scores rose from 167 to 188 from 2000 to 2009,<sup>129</sup> an improvement that corresponds to 2.1 grade levels. These results are impressive, especially when compared with the period from 1998 to 2000 in which fourth grade ELL scores decreased by 7 points.<sup>130</sup> In eighth grade reading, ELL scores remained fairly stable from 1998 to 2009, increasing slightly before 2002 and decreasing slightly after 2002.<sup>131</sup>

d. *Students with Disabilities*

NAEP data disaggregated by disability status is only available beginning in 1996. Consequently, as with ELL students, the lack of long-term trend data renders it difficult to analyze the effect that consequential accountability had on the scores of students with disabilities (“SWDs”). Still, there is some evidence that the rise of consequential accountability may have led to a rise in the reading scores of SWDs. In fourth grade reading, SWDs’ NAEP scores rose from 167 to 190 from 2000 to 2009, an improvement of 2.3 grade levels.<sup>132</sup> This growth narrowed the gap with non-SWDs by a substantial 14 points.<sup>133</sup> By comparison, from 1998 to 2000, the reading scores of fourth grade SWDs declined by 9 points.<sup>134</sup> Eighth grade NAEP reading

<sup>127</sup> The scores of ELLs are not disaggregated for the NAEP Long-Term Trend Assessment; for this reason, the discussion of ELL scores uses data from the standard NAEP Assessment rather than the Long-Term Trend Assessment. See Nat’l Ctr. for Educ. Statistics, U.S. Dep’t of Educ. Inst. Of Educ. Scis., *NAEP—2009 Reading: Grade 4 National Results*, THE NATION’S REPORT CARD, [http://www.nationsreportcard.gov/reading\\_2009/nat\\_g4.asp?subtab\\_id=Tab\\_7&tab\\_id=tab1](http://www.nationsreportcard.gov/reading_2009/nat_g4.asp?subtab_id=Tab_7&tab_id=tab1) (last visited Oct. 26, 2010).

<sup>128</sup> *Id.*

<sup>129</sup> *Id.*

<sup>130</sup> *Id.*

<sup>131</sup> From 1998 to 2009, eighth grade ELL reading scores rose by only one point, with scores rising slightly from 1998 to 2002 and declining somewhat from 2002 to 2009. *Id.*

<sup>132</sup> Nat’l Ctr. for Educ. Statistics, U.S. Dep’t of Educ. Inst. Of Educ. Scis., *NAEP—2009 Reading: Grade 4 National Results*, THE NATION’S REPORT CARD, [http://www.nationsreportcard.gov/reading\\_2009/nat\\_g4.asp?subtab\\_id=Tab\\_6&tab\\_id=tab1](http://www.nationsreportcard.gov/reading_2009/nat_g4.asp?subtab_id=Tab_6&tab_id=tab1) (last visited Oct. 26, 2010).

<sup>133</sup> *Id.*

<sup>134</sup> *Id.*

scores for SWDs increased at a relatively constant rate throughout this time period.<sup>135</sup>

#### 4. Effect on Mathematics Scores

##### a. General

As shown in Figure 2 below, average math scores for 9- and 13-year-olds on the NAEP Long-Term Trend Assessment have increased since 1999, although scores for 17-year-olds did not change significantly.<sup>136</sup> While 9-year-old average math scores only increased by 2 points from 1990 to 1999, they increased by 11 (13\*) points from 1999 to 2008.<sup>137</sup> The rate at which scores for 13-year-olds improved remained relatively constant before and after the widespread implementation of consequential accountability; from 1990 to 1999, average 13-year-old math scores increased by 6 points, whereas from 1999 to 2008, scores increased by 5 (7\*) points.<sup>138</sup> Score trends for 17-year-olds remained relatively flat, rising 3 points from 1990 to 1999 and decreasing by 2 (0\*) points from 1999 to 2008.<sup>139</sup>

FIGURE 2. CHANGES IN NAEP LONG-TERM TREND MATH SCORES.

	All Scores		Black Scores		Hispanic Scores	
	1990– 1999	1999– 2008	1990– 1999	1999– 2008	1990– 1999	1999– 2008
9-year-old	+2	+11	+3	+13	-1	+21
13-year-old	+6	+5	+2	+11	+4	+9
17-year-old	+3	-2	-6	+4	+9	0

##### b. Gains by Race

Black 9-year-olds made considerable gains in math from 1999 to 2008, improving their scores by 13 (16\*) points.<sup>140</sup> This gain was particularly encouraging considering that black improvement had stalled during the 1990s, with black 9-year-olds only improving 3 points in math from 1990 to 1999.<sup>141</sup> The white-black achievement gap in math, however, remained rela-

<sup>135</sup> Eighth grade reading scores for SWDs grew by 4 points from 1998 to 2002 and by 2 points from 2002 to 2009. Nat'l Ctr. for Educ. Statistics, U.S. Dep't of Educ. Inst. Of Educ. Scis., *NAEP—2009 Reading: Grade 8 National Results*, THE NATION'S REPORT CARD, [http://www.nationsreportcard.gov/reading\\_2009/nat\\_g8.asp?subtab\\_id=Tab\\_6&tab\\_id=tab1](http://www.nationsreportcard.gov/reading_2009/nat_g8.asp?subtab_id=Tab_6&tab_id=tab1) (last visited Oct. 26, 2010).

<sup>136</sup> RAMPEY ET AL., *supra* note 107, at 3.

<sup>137</sup> *Id.*

<sup>138</sup> *Id.*

<sup>139</sup> *Id.*

<sup>140</sup> *Id.* at 34.

<sup>141</sup> *Id.*

tively unchanged from 1990 to 2008, increasing by 1 point from 1990 to 1999 and decreasing by 2 points from 1999 to 2008.<sup>142</sup>

Black 13-year-olds also made significant gains in math from 1999 to 2008, gaining 11 (16\*) points, whereas from 1990 to 1999, this group only gained 2 points in math.<sup>143</sup> While the white-black achievement gap for 13-year-olds actually increased by 5 points from 1990 to 1999, this trend reversed during the period from 1999 to 2008, in which the achievement gap narrowed by 4 points.<sup>144</sup>

During the 1990s, the math scores of black 17-year-olds declined by 6 points.<sup>145</sup> From 1999 to 2008, black 17-year-olds recovered 4 of those lost points.<sup>146</sup> Similarly, the white-black math achievement gap for 17-year-olds widened by 10 points from 1990 to 1999, and then narrowed by 5 points from 1999 to 2008.<sup>147</sup>

Scores of Hispanic 9-year-olds in math went up from 213 to 234 from 1999 to 2008, a gain of 21 (22\*) points.<sup>148</sup> This roughly equates to an improvement of two grade levels, making it the largest gain recorded.<sup>149</sup> In contrast, the math scores of Hispanic 9-year-olds actually declined by 1 point from 1990 to 1999.<sup>150</sup> Moreover, during the 2000s, the white-Hispanic math achievement gap declined from 26 points in 1999 to only 16 points in 2008.<sup>151</sup> While there is still progress to make, Hispanic 9-year-olds are now performing about as well in math as whites were for much of the 1990s.<sup>152</sup> Hispanic 13-year-olds also made significant gains from 1999 to 2008, with scores rising from 259 to 268, a gain of 9 (10\*) points.<sup>153</sup> In comparison, from 1990 to 1999, Hispanic 13-year-olds only improved 4 points in math.<sup>154</sup> Because white 13-year-olds' math scores showed similar gains during the period 1990-2008, the Hispanic-white achievement gap for 13-year-olds remained relatively constant from 1990-2008.<sup>155</sup>

In sharp contrast to Hispanic 9- and 13-year-olds, the math scores of Hispanic 17-year-olds increased during the 1990s but decreased during the 2000s,<sup>156</sup> the period in which the rise in use of consequential accountability

<sup>142</sup> *Id.*

<sup>143</sup> *Id.* at 35.

<sup>144</sup> *Id.*

<sup>145</sup> *Id.*

<sup>146</sup> *Id.*

<sup>147</sup> *Id.*

<sup>148</sup> *Id.* at 36.

<sup>149</sup> In comparison, from 1973 to 1999, scores of Hispanic 9-year-olds only increased by 11 points. *Id.*

<sup>150</sup> *Id.*

<sup>151</sup> *Id.*

<sup>152</sup> In 2008, the average math score for Hispanic 9-year-olds was 234; the average score for white 9-year-olds was 235 in 1990 and 1992 and 237 in 1994 and 1996. *Id.*

<sup>153</sup> *Id.* at 37.

<sup>154</sup> Scores rose from 255 to 259. *Id.*

<sup>155</sup> The Hispanic-white math achievement gap for thirteen-year-olds increased by two points from 1990 to 1999 and decreased by one point from 1999 to 2008. *Id.*

<sup>156</sup> *Id.*

began to have widespread effects. The Hispanic-white achievement gap among 17-year-olds narrowed by 4 points from 1990 to 1999 and narrowed further by 1 point from 1999 to 2008.<sup>157</sup>

*c. English Language Learners*

English language learners made significant gains in math during the 2000s.<sup>158</sup> In fourth grade math, ELL students gained 19 points from 2000 to 2009, an improvement of roughly two grade levels.<sup>159</sup> By comparison, from 1996 to 2000, the scores of fourth grade ELL students decreased by 2 points.<sup>160</sup> In eighth grade math, ELL NAEP scores increased from 234 to 243 between 2000 and 2009, continuing the upward trend witnessed between 1996 and 2000.<sup>161</sup> While it is difficult to assess the long-term impact of consequential accountability on ELL math scores since ELL scores are only available beginning in 1996, the available data indicate that the widespread implementation of consequential accountability correlates with an increase in the achievement levels of ELL students.

*d. Students with Disabilities*

As with reading scores, it is difficult to assess the impact of consequential accountability on the math scores of students with disabilities because of the lack of long-term data.<sup>162</sup> An analysis of the available data, however, indicates that the use of consequential accountability systems correlates with positive gains in the achievement levels of SWDs. In fourth grade math, for example, the NAEP scores of SWDs grew from 198 to 221 from 2000 to 2009, a remarkable improvement of roughly 2.3 grade levels.<sup>163</sup> During this time period, non-SWD scores grew from 228 to 242, and thus the gap between SWD and non-SWD students narrowed by 9 points from 2000 to 2009.<sup>164</sup> From 1996 to 2000, the math scores of fourth grade SWDs

<sup>157</sup> *Id.*

<sup>158</sup> The source of ELL math data is the standard NAEP assessment, which only has data for ELL students beginning in 1996. Nat'l Ctr. for Educ. Statistics, U.S. Dep't of Educ. Inst. Of Educ. Scis., *NAEP—Mathematics 2009: Grade 4 National Results*, THE NATION'S REPORT CARD, [http://www.nationsreportcard.gov/math\\_2009/gr4\\_national.asp?subtab\\_id=Tab\\_7&tab\\_id=tab1](http://www.nationsreportcard.gov/math_2009/gr4_national.asp?subtab_id=Tab_7&tab_id=tab1) (last visited Oct. 26, 2010).

<sup>159</sup> *Id.*

<sup>160</sup> *Id.*

<sup>161</sup> From 1996 to 2000, eighth grade ELL scores in math increased by 8 points. Nat'l Ctr. for Educ. Statistics, U.S. Dep't of Educ. Inst. Of Educ. Scis., *NAEP—Mathematics 2009: Grade 8 National Results*, THE NATION'S REPORT CARD, [http://www.nationsreportcard.gov/math\\_2009/gr8\\_national.asp?subtab\\_id=Tab\\_7&tab\\_id=tab1](http://www.nationsreportcard.gov/math_2009/gr8_national.asp?subtab_id=Tab_7&tab_id=tab1) (last visited Oct. 26, 2010).

<sup>162</sup> Data on the math scores of students with disabilities are available beginning in 1996. Nat'l Ctr. for Educ. Statistics, U.S. Dep't of Educ. Inst. Of Educ. Scis., *NAEP—Mathematics 2009: Grade 4 National Results*, THE NATION'S REPORT CARD, [http://www.nationsreportcard.gov/math\\_2009/gr4\\_national.asp?subtab\\_id=Tab\\_6&tab\\_id=tab1](http://www.nationsreportcard.gov/math_2009/gr4_national.asp?subtab_id=Tab_6&tab_id=tab1) (last visited Oct. 26, 2010).

<sup>163</sup> *Id.*

<sup>164</sup> *Id.*

decreased by 6 points.<sup>165</sup> In eighth grade math, the NAEP scores of SWDs increased from 230 to 249 from 2000 to 2009.<sup>166</sup> By comparison, from 1996 to 2000 eighth grade math scores for SWDs decreased by 1 point.<sup>167</sup> Thus, while there is insufficient data to do a long-term trend analysis of the scores of SWDs, the available NAEP results appear to support the idea that consequential accountability may have had a positive impact on the achievement levels of SWDs.

### 5. *Summary of Post-1999 Data*

Data from the NAEP Long-Term Trend Assessment confirm that many groups of students made large positive gains in achievement in the 2000s, the period in which one would expect to see the widespread effects of consequential accountability. As displayed in Figure 1 and Figure 2 above, the results at the fourth grade level are most impressive, with all groups of students making significantly larger gains in achievement in the 2000s than they made during the 1990s. The results at the eighth grade level show impressive gains in math, especially for blacks and Hispanics. The data also show that black students made impressive gains across the board in the 2000s compared to their score trends in the 1990s, a phenomenon which may be due, at least in part, to NCLB's data disaggregation requirement.

Although the main NAEP assessment does not provide enough long-term data to analyze accurately the impact of consequential accountability on the achievement levels of ELLs and SWDs, the available data are largely positive: at the fourth grade level, all ELLs and SWDs made significant gains in the 2000s compared to their prior score trends. At the eighth grade level, score trends remained relatively constant, with the exception that SWDs made significant gains in eighth grade math. While more research is needed to determine whether the results mentioned above hold when controlling for other factors, the score trends support the proposition that the rise of consequential accountability was likely a major cause of substantial gains in student achievement.

### 6. *A Note on High School Performance*

NAEP data clearly show stagnation in the scores for twelfth graders. While there are no proven explanations of this phenomenon, and further research is essential, it merits at least a brief consideration of possible hypotheses. There are both intrinsic and extrinsic problems associated with policy

---

<sup>165</sup> *Id.*

<sup>166</sup> Nat'l Ctr. for Educ. Statistics, U.S. Dep't of Educ. Inst. Of Educ. Scis., *NAEP—Mathematics 2009: Grade 8 National Results*, THE NATION'S REPORT CARD, [http://www.nationsreportcard.gov/math\\_2009/gr8\\_national.asp?subtab\\_id=Tab\\_6&tab\\_id=tab1](http://www.nationsreportcard.gov/math_2009/gr8_national.asp?subtab_id=Tab_6&tab_id=tab1) (last visited Oct. 26, 2010).

<sup>167</sup> *Id.*

and practice that may be the cause or share the cause in twelfth grade NAEP stagnation. First, most consequential accountability systems in place today do not “pinch,” or “pinch” very little beyond middle school. There is no clear pattern of testing at the high school level for accountability purposes and no clear set of standards that apply at the high school level. Intrinsically, by focusing accountability on education through the eighth grade, and largely abandoning accountability thereafter, this policy forgoes the benefits of accountability for students who have passed the eighth grade. Put simply, the incentive structure that has proved effective for younger students has not yet been effected (largely) for those in high school. As more states extend consequential accountability through high school, this hypothesis will be able to be tested.

Extrinsic from the existence or non-existence of accountability in post-eighth grade education, there may be a problem with testing twelfth graders as a matter of practicality that would be difficult to correct. The NAEP test, as any 17-year-old student would know, will affect few, if any, of the choices that are important to the students personally, such as graduation, college, or a career. Hence, there may be a striking motivational difference between the twelfth grade NAEP test-taker and the more compliant fourth or eighth grade NAEP test-taker that could dilute the meaningfulness of the former data.

Finally, standardization is much more characteristic of elementary and middle schools than high schools. In other words, most can agree what basic skills and knowledge a younger student should master from grade to grade. It is much more difficult to develop common standards where there are varying curricula more specifically tailored to different students in high school. Setting standards, measuring success toward learning to those standards, and imposing consequences—the essential elements of consequential accountability—have not yet been implemented generally across the nation’s high schools. States are increasingly trying to address this problem. Time will tell if these efforts will improve NAEP results for 17-year-olds.

#### IV. CHARACTERISTICS OF CONSEQUENTIAL ACCOUNTABILITY SYSTEMS

This Article has shown in Part III that consequential accountability appears a likely cause of gains in student achievement.<sup>168</sup> The question then becomes, what can one do with that knowledge? In the following Parts, the authors hope to provide the reader with the tools necessary to analyze current proposals for the reform of NCLB. In order to evaluate future proposals

---

<sup>168</sup> Data from the 1990s strongly support the proposition that consequential accountability leads to significant positive gains in student achievement. See *supra* Part III.A. More research is needed to assess data from the 2000s; however, a preliminary examination of the data supports the finding that the growth of the consequential accountability movement likely had positive effects on student achievement in the 2000s, especially among certain subgroups and at the fourth grade level. See *supra* Part III.B.



for the reform of consequential accountability systems, however, one must understand what the consequential accountability systems of the past looked like and how they differed from one another. In this Part, this Article will first discuss the elements of two state accountability systems that were particularly successful in the 1990s before turning to a discussion of NCLB's characteristics.

### A. State Accountability Systems

As of 1999, the two states that practiced consequential accountability most effectively were Texas and North Carolina. These two states created accountability systems that identified both high and low performers and attached rewards and consequences to performance.<sup>169</sup> As noted above, Texas and North Carolina also experienced the greatest gains in student test scores in the nation during the 1990s.<sup>170</sup> Significantly, researchers note that the gains experienced by Texas and North Carolina in the 1990s were the result of policies put into place during the late 1980s and early 1990s, particularly policies that: (1) developed systems of accountability that measured student performance against state standards; (2) created computerized feedback systems; and (3) attached consequences to performance.<sup>171</sup> This Article will examine the accountability systems of both Texas and North Carolina in an effort to learn from that success.

The story of consequential accountability in Texas is particularly important since Texas is often considered the "birthplace of NCLB."<sup>172</sup> The Texas educational accountability system was created in 1993 with the passage of the politically controversial Senate Bill 7, a bill whose principal purpose was to overhaul school funding within the state.<sup>173</sup> The bill, which built on existing standards and testing mechanisms,<sup>174</sup> called for the State Board of Education to develop standards that met the goals of the bill,<sup>175</sup> establish state-

---

<sup>169</sup> Craig D. Jerald & Ulrich Boser, *Who's Holding Schools Responsible?*, 18 EDUC. WEEK 81 (1999), available at <http://www.edcounts.org/archive/sreports/qc99/ac/tables/ac-c1.htm>. While many states implemented consequential accountability systems in the 1990s, Texas and North Carolina were unique in that their statutes attached consequences to both high and low performance. *Id.* However, according to David Dunn, "while these rewards were authorized, they have not been funded on a continuing basis." Telephone Interview with David Dunn, Exec. Dir., Tex. Charter Sch. Ass'n (Nov. 12, 2010).

<sup>170</sup> David Grissmer & Ann Flanagan, *Exploring Rapid Achievement Gains in North Carolina and Texas*, NAT'L EDUC. GOALS PANEL, 6-7 (1998), available at <http://govinfo.library.unt.edu/negp/reports/grissmer.pdf>.

<sup>171</sup> *Id.*

<sup>172</sup> Sarah W. Nelson et al., *Fulfilling the Promise of Educational Accountability*, 88 PHI DELTA KAPPAN 702, 702 (2007).

<sup>173</sup> Tex. S.B. 7, 73d Leg., R.S., available at <http://www.legis.state.tx.us/tlodocs/73R/billtext/html/SB00007F.html> (subsequently enacted as Act of May 28, 1993, 73d Leg., 1993 Tex. Gen. Laws 1479).

<sup>174</sup> Grissmer & Flanagan, *supra* note 170, at 29.

<sup>175</sup> Tex. S.B. 7 § 35.021.

wide testing against those standards,<sup>176</sup> and establish minimum proficiency levels for “satisfactory performance.”<sup>177</sup> The bill also required assessment of the performance of schools and districts by both comparing performance against set state standards and by examining year-to-year improvements by schools and districts.<sup>178</sup>

The Texas bill contained several groundbreaking features. One simple, yet noteworthy, idea underlying the accountability system was the principle that “all children can learn.”<sup>179</sup> Jerald notes that “[i]n many states, efforts to raise standards have unearthed an ugly reality beneath the rhetoric that ‘all children can learn’”—states often set lower standards for certain groups of students or blame poor test performance on challenges such as poverty or family factors.<sup>180</sup> In contrast, the Texas accountability system sought to increase transparency and focus on subgroups by requiring disaggregation of test scores by race, ethnicity, gender, and socioeconomic status.<sup>181</sup> Texas also supported the idea that “all children can learn” by requiring each subgroup to meet the same standards,<sup>182</sup> allowing students with serious disabilities as the only exception.<sup>183</sup>

These features—the disaggregation of data and the requirement that all subgroups be held to the same standards—while groundbreaking at the time, eventually became a national norm when they were included as key features of NCLB in 2001. By focusing on the outcomes for specific subgroups, Texas designed its accountability system to overcome many of the inequalities that persisted in the educational system.<sup>184</sup>

Another noteworthy feature of the Texas consequential accountability system relates to how Texas set the proficiency levels used to measure schools. Rather than setting high performance standards at the start, “Texas intentionally set its bar just above where the system had been performing up to 1993 and then raised it incrementally [over] the course of the decade.”<sup>185</sup> This strategic decision, which allowed schools to work toward *attainable* goals as performance standards were gradually raised over time, appears to have been successful at raising achievement in Texas, even as standards have

<sup>176</sup> Tex. S.B. 7 § 35.022.

<sup>177</sup> Tex. S.B. 7 § 35.024.

<sup>178</sup> Tex. S.B. 7 § 35.041(b).

<sup>179</sup> CRAIG D. JERALD, *Real Results, Remaining Challenges: The Story of Texas Education Reform*, in THE BUSINESS ROUNDTABLE 3 (2001), available at <http://eric.ed.gov/ERICWebPortal/contentdelivery/servlet/ERICServlet?accno=ED459191>.

<sup>180</sup> *Id.*

<sup>181</sup> Tex. S.B. 7 § 35.041(b).

<sup>182</sup> JERALD, *supra* note 179, at 3.

<sup>183</sup> See Tex. S.B. 7 § 35.027(a) (directing that “[a]ny student who has a physical or mental impairment or a learning disability that prevents the student from mastering the competencies which the academic skills assessment instruments are designed to measure may be exempted from the requirements of this subchapter”).

<sup>184</sup> Nelson et al., *supra* note 172, at 703.

<sup>185</sup> JERALD, *supra* note 179, at 3.

been raised.<sup>186</sup> While some observers criticized Texas for setting low standards initially, Texas raised those standards significantly over the first five years of the program.<sup>187</sup> As Jerald put it, “[l]ike a magnet that can pull a metal object only if one positions it close enough and then moves it slowly away, the Texas reforms met schools where they were and pulled them into higher performance over time.”<sup>188</sup>

Also remarkable were the consequences Texas attached to these assessments. The Texas system incorporated several features to provide incentives for schools to improve their performance. For rewards, the Texas system was designed to provide financial benefits for those schools that “demonstrate progress or success in achieving the education goals of the state” as enumerated in the statute.<sup>189</sup> For penalties, Texas provided for a number of possible sanctions, including the development of student achievement improvement plans,<sup>190</sup> the use of public hearings,<sup>191</sup> the appointment of a school monitor,<sup>192</sup> and state takeover of poorly performing schools.<sup>193</sup>

Some researchers have opined, however, that these accountability mechanisms were weak, in part because the sanctions did not go into effect automatically.<sup>194</sup> Additionally, Texas has used its most severe sanctions—supervision by a monitor or master and reconstitution—only sparingly.<sup>195</sup> Instead, Texas primarily focused on shining a “spotlight” on poorly performing schools in an effort both to “shame” them into better performance and to cause districts to bring improvements to bear using school report

<sup>186</sup> ACHIEVE, INC., AIMING HIGHER: MEETING THE CHALLENGES OF EDUCATION REFORM IN TEXAS 19 (2002), [http://www.achieve.org/files/Aiming\\_Higher\\_Texas.pdf](http://www.achieve.org/files/Aiming_Higher_Texas.pdf). For example, Jerald notes that in the years following the implementation of Texas’s accountability system, “Texas [saw] dramatic increases in passing rates across all grade levels and subjects on the state’s Texas Assessment of Academic Skills (TAAS) exams, coupled with equally dramatic decreases in racial and socioeconomic gaps on those tests.” JERALD, *supra* note 179, at 6.

<sup>187</sup> JERALD, *supra* note 179, at 5. For example, for the “Academically Acceptable” rating, minimum proficiency levels rose by 5% per year from 1995 to 2000. *Id.*

<sup>188</sup> *Id.* at 4.

<sup>189</sup> Tex. S.B. 7 § 35.081, 73d Leg., R.S., available at <http://www.legis.state.tx.us/tlodocs/73R/billtext/html/SB00007F.html> (subsequently enacted as Act of May 28, 1993, 73d Leg., 1993 Tex. Gen. Laws 1479). However, according to David Dunn, “while these rewards were authorized, they have not been funded on a continuing basis.” Telephone Interview with David Dunn, Exec. Dir., Tex. Charter Sch. Ass’n (Nov. 12, 2010).

<sup>190</sup> Tex. S.B. 7 § 35.121(a)(3). This sanction allows the commissioner of education to “order the preparation of a student achievement improvement plan that addresses each academic excellence indicator for which the district’s performance is unacceptable” and order the implementation of the plan. *Id.*

<sup>191</sup> Tex. S.B. 7 § 35.121(a)(2). Public hearings function as a sanction because they draw attention to the fact that the school has not met established standards.

<sup>192</sup> Tex. S.B. 7 § 35.121(a)(6)-(7); Heinrich Mintrop & Tina Trujillo, *Corrective Action in Low Performing Schools: Lessons for NCLB Implementation from First-generation Accountability Systems*, 13 EDUC. POL’Y ANALYSIS ARCHIVES 1, 5 (Dec. 1, 2005), available at <http://epaa.asu.edu/ojs/article/viewFile/153/279>.

<sup>193</sup> Tex. S.B. 7 § 35.121(a)(8)-(9); Grissmer & Flanagan, *supra* note 170, at 29.

<sup>194</sup> ACHIEVE, INC., *supra* note 186, at 25.

<sup>195</sup> Mintrop & Trujillo, *supra* note 192, at 1, 6.

cards, public announcements about performance and interventions.<sup>196</sup> Thus, the effectiveness of Texas's accountability system appears due, in part, to the impact that the *threat* of sanctions had on creating positive changes in the operation of the schools. As such, Texas's ability to identify poorly performing schools and the availability of harsher sanctions allowed the state to put strong pressure on poorly performing schools and motivate them to higher levels of performance.<sup>197</sup>

Like Texas, North Carolina implemented a comprehensive accountability system during the 1990s. Standards and assessments were first introduced in North Carolina in 1992 and 1993<sup>198</sup> and a comprehensive education reform bill was signed into law in 1996.<sup>199</sup> The bill, which the North Carolina State Board of Education designed, was called the "ABCs of Public Education" ("ABCs"), an acronym that stands for "Accountability; teaching the Basics of reading, writing, and mathematics; and increasing Control of schools at the local level" by the Board.<sup>200</sup> North Carolina first implemented the ABCs law during the 1996-1997 school year.<sup>201</sup>

The ABCs law contained five key parts, which the bill outlines as:

the State Board of Education shall adopt guidelines . . . to: (1) Assist local boards and schools in the development and implementation of school-based management . . . (2) Recognize the schools that meet or exceed their goals. (3) Identify low-performing schools . . . and create assistance teams that the Board may assign to schools identified as low-performing . . . (4) Enable assistance teams to make appropriate recommendations [to the school] . . . (5) Establish a process to resolve disputes between local boards and schools in the development and implementation of school improvement plans . . .<sup>202</sup>

According to Jay Robinson, the Chairman of the North Carolina State Board of Education at the time North Carolina adopted the ABCs law, the purpose of the law was to "[make] sure your child gets at least a year's worth of growth for a year's worth of school."<sup>203</sup>

Upon first glance, the Texas and North Carolina accountability laws are similar in that both seek to ensure that all children make appropriate pro-

<sup>196</sup> ACHIEVE, INC., *supra* note 186, at 25; Mintrop & Trujillo, *supra* note 192, at 1, 6.

<sup>197</sup> Mintrop & Trujillo, *supra* note 192, at 1, 12.

<sup>198</sup> Grissmer & Flanagan, *supra* note 170, at 32.

<sup>199</sup> School-Based Management and Accountability Program, N.C. GEN. STAT. ANN. § 115c-105.20 (2010).

<sup>200</sup> ABCs OF PUBLIC EDUCATION: A JOURNEY TOWARD EXCELLENCE 3 (North Carolina Board of Education) (1999), available at <http://eric.ed.gov/ERICWebPortal/contentdelivery/servlet/ERICServlet?accno=ED469452> [hereinafter ABCs OF PUBLIC EDUCATION]. Perhaps the Board of Education can be forgiven for using the rather attenuated shortened name. *Id.*

<sup>201</sup> *Id.* at 2.

<sup>202</sup> N.C. GEN. STAT. ANN. § 115c-105.20(b).

<sup>203</sup> ABCs OF PUBLIC EDUCATION, *supra* note 200, at 3.

gress each year in school through testing accountability. The most readily apparent difference is that in addition to measuring student proficiency against predetermined state standards, the North Carolina bill also allowed the state to use student growth measures in its accountability system.<sup>204</sup> Thus, North Carolina measured schools' performance against both the common benchmarks set for all schools and the growth of their students' scores over the course of the school year.<sup>205</sup> Under this approach, schools were categorized based on both whether they met their expected growth each year *and* the percentage of their students that scored at or above a set achievement level.<sup>206</sup> Another difference between the Texas and North Carolina laws is that despite using strong rhetoric about helping the individual student,<sup>207</sup> North Carolina did not require disaggregation of student test scores by subgroups as Texas did.<sup>208</sup>

In order to hold schools accountable, the ABCs law, like the Texas law, provides both rewards and sanctions for schools based on their performance on state tests.<sup>209</sup> The state categorizes schools based on the average gains of its students compared to students' expected growth, as calculated by the state.<sup>210</sup> The state then provides rewards through monetary bonuses to teachers and staff in all "exemplary schools," which are schools that exceed their performance targets by at least ten percent.<sup>211</sup>

As for sanctions, North Carolina took a fairly innovative approach to accountability in designing the ABCs law<sup>212</sup>: rather than "punishing" schools that did not meet their performance targets, the state under the original ABCs law sent in "Assistance Teams" that made recommendations to the school for how to improve student achievement, worked with the local district to implement a plan for improving student achievement, and monitored the school's progress under the plan.<sup>213</sup> In practice, these assistance teams were typically made up of an administrator and three or four teachers

<sup>204</sup> N.C. GEN. STAT. ANN. § 115C-105.27(a).

<sup>205</sup> COLUMBUS CNTY. SCHS., OVERVIEW OF THE ABCs OF PUBLIC EDUCATION 1 (2003), <http://www.columbus.k12.nc.us/hms/FAQ%20documents/abcs.pdf>.

<sup>206</sup> *Id.*

<sup>207</sup> N.C. GEN. STAT. ANN. § 115C-105.3.

<sup>208</sup> The disaggregation of data would have allowed the state to measure the progress of subgroups of students, thereby identifying low-performing groups and setting goals for the progress of those groups.

<sup>209</sup> Kathleen Kennedy Manzo, *Policy Update: North Carolina Seeing a Payoff-Quality Counts '99*, EDUC. WEEK ON THE WEB, <http://www.edcounts.org/archive/sreports/qc99/states/policy/nc-up.htm> (last visited Oct. 26, 2010).

<sup>210</sup> Helen F. Ladd & Arnaldo Zelli, *School-Based Accountability in North Carolina: The Responses of School Principals*, 38 EDUC. ADMIN. Q. 494, 499 (2002), available at <http://eaq.sagepub.com/content/38/4/494.full.pdf+html>.

<sup>211</sup> Manzo, *supra* note 209; Ladd & Zelli, *supra* note 210, at 500.

<sup>212</sup> COLUMBUS CNTY. SCHS., *supra* note 205, at 2.

<sup>213</sup> ABCs of Public Education, ch. 716, 1995 N.C. Sess. Laws 352, 368-69. Since the implementation of the ABCs law, North Carolina has made some changes to its accountability sanctions. Currently, if a school is identified as "low performing" under the ABCs law, it becomes eligible to receive assistance from the District and School Transformation Division.

with experience in the grade range of the particular school. The teams, which received extensive training, worked with schools on a daily basis for one year once a school was identified as having not met its performance target.<sup>214</sup> This structure in practice meant that North Carolina's "sanctions" were actually a tool to help schools get on the right track and improve their performance.

### B. *The No Child Left Behind Act*

The No Child Left Behind Act of 2001 marked a historic shift in federal education policy.<sup>215</sup> Gone were the days of limited or no federal involvement in education. NCLB set the ambitious goal of having all children achieve proficiency in reading and math by 2014 and created federal expectations of state accountability systems in an effort to achieve that goal.<sup>216</sup>

The major hallmark of NCLB is its focus on improving student achievement for *all* students, with a particularly strong focus on those students who were traditionally disadvantaged or overlooked.<sup>217</sup> Like the Texas accountability system, NCLB requires that data be disaggregated for various subgroups, including economically disadvantaged students, students with disabilities, students with limited English proficiency, major racial and ethnic groups, and gender groups.<sup>218</sup> This feature created a positive shift in public education; as then-Secretary of Education Margaret Spellings put it, "[f]or the first time ever, we are looking ourselves in the mirror and holding ourselves accountable for educating every child. That means all children, no matter their race or income level or zip code."<sup>219</sup>

The disaggregation requirement, when coupled with the requirement that decision-makers consider key subgroups in determining whether schools make Adequate Yearly Progress, is perhaps NCLB's most laudable feature. In the years following the passage of NCLB, the scores of many of the subgroups that NCLB targeted experienced dramatic gains.<sup>220</sup> While more research is needed to determine whether these gains specifically can be attributed to NCLB, the NAEP data provide strong evidence to support the hypothesis that NCLB's disaggregation requirement had a positive impact generally on the scores of the subgroups that it targeted.

<sup>214</sup> Mintrop & Trujillo, *supra* note 192, at 15.

<sup>215</sup> No Child Left Behind Act of 2001, 20 U.S.C. §§ 6301-7941 (2006).

<sup>216</sup> Dee & Jacob, *supra* note 95, at 2.

<sup>217</sup> Caroline M. Hoxby, *Inadequate Yearly Progress: Unlocking the Secrets of NCLB*, 5 EDUC. NEXT 46, 48 (2005), available at [http://educationnext.org/files/ednext20053\\_46.pdf](http://educationnext.org/files/ednext20053_46.pdf).

<sup>218</sup> Richard Wenning et al., *No Child Left Behind: Testing, Reporting, and Accountability*, EDUC. RESOURCES INFO. CTR. DIG., Aug. 2003, at 1, 3, available at <http://eric.ed.gov/ERICWebPortal/contentdelivery/servlet/ERICServlet?accno=ED480994>.

<sup>219</sup> U.S. DEP'T OF EDUC., HOW *NO CHILD LEFT BEHIND* BENEFITS AFRICAN AMERICANS, available at <http://www2.ed.gov/nclb/accountability/achieve/nclb-aa.html>.

<sup>220</sup> See *supra* Part III.B.

Another notable feature of NCLB is its emphasis on state autonomy. Rather than setting common standards against which all states will be measured, NCLB allows each state to develop its own standards and testing mechanisms.<sup>221</sup> While the creation of common standards would likely not have been politically feasible at the time of NCLB's passage in 2001, the wide variation of standards across states is often cited as one of the greatest weaknesses in the design of NCLB.<sup>222</sup> However, under NCLB, all states were required to participate in the NAEP assessment and evaluators began giving the NAEP test every two years, allowing for better comparisons between states. This requirement drove the movement toward common standards by allowing for a more robust comparison of state standards and thereby forcing a consideration of the comparability of standards across states.

Despite its focus on state autonomy, NCLB is rigid in its requirement that states design their accountability systems with the goal of all subgroups achieving proficiency by 2014. While states are still left to create their own standards and assessments, NCLB's 2014 proficiency deadline limits state flexibility since states cannot design their accountability systems around an alternative goal or deadline. This component of NCLB sharply diverges from the Texas system in which the state chose to increase proficiency scores over time at what was intended to be an attainable gradual rate. NCLB also differs from the North Carolina approach, which allows the state to make individualized determinations about the progress that each school can and should make from year to year. In striking contrast to the North Carolina system, NCLB requires states to set statewide proficiency targets that all schools, along with each major subgroup within each school, must meet in order to make AYP.<sup>223</sup>

A key feature of any consequential accountability system is the consequences that it imposes for poor performance. Unlike the Texas and North Carolina consequential accountability systems, NCLB solely focused on sanctions for poor performance, rather than providing incentives for exemplary performance.<sup>224</sup> Under NCLB, schools that do not meet AYP are sub-

---

<sup>221</sup> Hoxby, *supra* note 217, at 49.

<sup>222</sup> *Id.* at 48.

<sup>223</sup> Jessica Flynn, Note, *No Child Left Behind Gets a Tip from Tort Theory: Protecting Responsible Schools Against Undeserved Sanctions*, 31 T. JEFFERSON L. REV. 157, 166 (2008).

<sup>224</sup> In fact, President George W. Bush's administration's original NCLB proposal did call for rewards, although this provision was ultimately removed because it faced significant opposition in Congress from both the left and the right. NO CHILD LEFT BEHIND, OFFICE OF PRESIDENT GEORGE W. BUSH 9 (2001), available at <http://www2.ed.gov/nclb/overview/intro/presidentplan/proposal.pdf> (noting that the Bush administration proposed that "[s]chools and states that make significant progress in closing the achievement gap will be honored with awards from a 'No Child Left Behind' school bonus fund and an 'Achievement in Education' state bonus fund").

ject to escalating interventions for each year that they continue to fail to make AYP.<sup>225</sup>

Under NCLB, a school is first identified as in need of improvement if it fails to make AYP for two consecutive years.<sup>226</sup> Once identified as “in need of improvement,” in the following year, which is known as “School Improvement Year 1,” the school district must allow students in the “failing” school to transfer to other schools, provide technical assistance to the school, and develop a school improvement plan.<sup>227</sup> If the school fails to meet AYP for the third year in a row, the district must provide supplemental education services to low-income students, as well as continue to provide technical assistance to the school during the next year, known as “School Improvement Year 2.”<sup>228</sup> If a school fails to make AYP for four consecutive years, the school must take corrective action during the fifth year.<sup>229</sup> Corrective action can include replacing relevant school staff,<sup>230</sup> implementing a new curriculum, decreasing management authority in the school, appointing outside experts to advise the school, extending the school year or school day, or restructuring the school.<sup>231</sup> If after one year of corrective action, the school still fails to make AYP, meaning that it has failed to make AYP for five consecutive years, it must develop a restructuring plan, which goes into effect the following year if it again fails to make AYP.<sup>232</sup>

Districts can choose from several options for restructuring: (1) close and reopen a school as a charter; (2) replace school staff; (3) turn over a school’s governance to the state; (4) contract with a private management company to operate the school; or (5) “any other major restructuring of the school’s governance designed to produce major reform.”<sup>233</sup> The most popular of these options has been “any other major restructuring,” an option that is fairly flexible and may result in much less aggressive reform than might be imagined.<sup>234</sup> For this reason, NCLB’s ultimate sanctions may be less stringent than one would expect.

The design of NCLB allows for the use of fairly severe sanctions relatively quickly.<sup>235</sup> It incorporates the Texas idea of shining a “spotlight”<sup>236</sup> on

<sup>225</sup> See *supra* Part II.B.

<sup>226</sup> No Child Left Behind Act, 20 U.S.C. § 6316(b)(1)(A) (2006).

<sup>227</sup> TEX. ASSOC. OF SCH. BDS., A COMPARISON OF THE NCLB AND TEXAS ACCOUNTABILITY SYSTEMS 6 (2008), [http://www.tasb.org/legislative/documents/account\\_diff1008.pdf](http://www.tasb.org/legislative/documents/account_diff1008.pdf).

<sup>228</sup> *Id.* at 7.

<sup>229</sup> *Id.*

<sup>230</sup> The statute does not define “relevant school staff.” It merely states that a school may “replace the school staff who are *relevant to the failure to make adequate yearly progress.*” 20 U.S.C. § 6316(b)(7)(C)(iv)(I).

<sup>231</sup> 20 U.S.C. § 6316(b)(7)(C)(iv)(VI).

<sup>232</sup> 20 U.S.C. § 6316(b)(8)(B).

<sup>233</sup> Mead, *supra* note 51, at 53.

<sup>234</sup> *Id.* at 53-54. For example, Mead notes that in Michigan, schools can “restructure” by simply hiring a state-trained “coach” who will design and implement a school improvement plan. *Id.* at 54.

<sup>235</sup> Mintrop & Trujillo, *supra* note 192, at 7.

<sup>236</sup> ACHIEVE, INC., *supra* note 186, at 25.



low-performing schools, as well as bits and pieces of the North Carolina approach, which focuses on providing assistance to struggling schools. But NCLB also may allow for too wide a variation in its sanctions: the threat of “any other major restructuring” for a school that fails to make AYP for five years in a row may not be sufficiently aggressive to address the issues that underlie the school’s inability to make AYP.<sup>237</sup>

Another potential problem with NCLB is that the spotlight that it shines on schools in need of improvement may become too wide in scope as it captures increasing numbers of schools who cannot meet proficiency targets for all subgroups by 2014, rather than illuminating only those schools that are most in need of improvement. Consequently, the law may lose some of its effectiveness as the 2014 proficiency target approaches. This Article will discuss these and other issues in the Parts that follow as it addresses some of the problems of consequential accountability systems and makes recommendations for the future reform of NCLB.

## V. PROBLEMS AND NEGATIVE EXTERNALITIES OF CONSEQUENTIAL ACCOUNTABILITY SYSTEMS AND NCLB

In the previous Part, the Article outlines the forms that several consequential accountability systems have taken, as well as some of their particularly successful or innovative features and how those features compare to one another. Any mature analysis into the effect of consequential accountability generally, or NCLB’s application of its principles specifically, will also need to account for the various ill effects (some real, some imagined) associated with consequential accountability or NCLB. Since imperfect people make imperfect laws, one aim of public policy should be to perfect the work that has already been done. To aid this Article’s final analysis of the Obama administration’s proposed modifications to NCLB, this Part will discuss the various problems and negative externalities commonly associated with consequential accountability systems. Some of these problems may be applicable to *any* consequential accountability system (e.g., alleged “teaching to the test” or curriculum narrowing) while others are specific to NCLB (e.g., alleged downward pressure on standards).

### A. Problems with Performance Standards and Content Standards

Perhaps the most written-about problem, specific to NCLB, is the downward pressure that critics allege the law puts on performance and content standards.<sup>238</sup> While NCLB provides that schools must meet AYP and

---

<sup>237</sup> Mead, *supra* note 51, at 53.

<sup>238</sup> To be clear, there is a difference between “content standards” and “performance standards.” The latter refers to the standards used to determine proficiency on tests whereas the former are used generally to describe the knowledge and skills that the state intends that its students attain. This Part concerns issues around performance standards.

must achieve proficiency for all subgroups by 2014, each state defines both AYP and proficiency. Consequently, a state can increase its proportion of schools meeting AYP by lowering the “proficiency” bar; increasing standards has the opposite effect. Peterson describes this issue in dramatic terms, stating:

In Massachusetts, for example, 43 percent of the students failed to make AYP, despite the fact that the state has the highest-performing students in the country. Why? Because Massachusetts has one of the highest standards in the country, a standard as high as the one NAEP uses. Conversely, only 7 percent of the schools in Tennessee are failing, though the state ranks near the bottom in terms of school performance. Why? Because Tennessee has one of the lowest operational definitions of proficiency in the country.<sup>239</sup>

This phenomenon causes two potential problems. First, comparing states becomes increasingly difficult as their respective standards diverge and the *actual* (rather than legal) significance of whether a school meets AYP becomes severely diluted in some states. There is currently wide variation in standards across states and wide variation in the rigor of state tests,<sup>240</sup> meaning that some states set “proficiency” at a much lower bar.<sup>241</sup> As Peterson and Hess put it rather cynically, “Johnny can’t read . . . in South Carolina. But if his folks move to Texas, he’ll be reading up a storm.”<sup>242</sup>

Second, states may have a strong disincentive against setting higher standards as higher standards will likely lead to a greater number of their schools being labeled as “failing,” a label likely of great concern to schools.<sup>243</sup>

Critics diverge on which of the two problems to highlight.<sup>244</sup> In truth, only the latter problem should matter in discussing NCLB reforms. Assume

<sup>239</sup> Paul E. Peterson, *A Lens That Distorts*, 7 *EDUC. NEXT* 46, 49 (2007), available at [http://educationnext.org/files/ednext\\_20074\\_46.pdf](http://educationnext.org/files/ednext_20074_46.pdf).

<sup>240</sup> This is not to say that this variation is related to NCLB—indeed, there was wide variation in standards long before NCLB became effective—the only point here pertains to the interstate comparability of NCLB’s AYP bifurcation.

<sup>241</sup> Again, this dilution could affect either performance standards or content standards. Paul E. Peterson & Frederick M. Hess, *Johnny Can Read . . . : Assessing the Rigor of State Assessment Systems*, 5 *EDUC. NEXT* 52, 52-53 (2005), available at [http://educationnext.org/files/ednext20053\\_52.pdf](http://educationnext.org/files/ednext20053_52.pdf).

<sup>242</sup> *Id.*

<sup>243</sup> See Chester E. Finn, Jr. & Michael Petrilli, *A New New Federalism: The case for national standards and tests*, 6 *EDUC. NEXT* 48, 50-51 (2006), available at [http://educationnext.org/files/ednext20064\\_48.pdf](http://educationnext.org/files/ednext20064_48.pdf) (making the point that since federal funding will come regardless of AYP, schools are probably more concerned with “irate local superintendents and school board members who don’t want NCLB to shine a harsh light on their schools’ shortcomings”). David Figlio and Cecilia Elena Rouse also provided a study indicating that the threat of failure “stigma” has been an effective motivator for increasing achievement in Florida schools. David N. Figlio & Cecilia Elena Rouse, *Do Accountability and Voucher Threats Improve Low-Performing Schools?*, 90 *J. OF PUB. ECON.* 239, 239-55 (2006).

<sup>244</sup> Compare, e.g., Paul E. Peterson & Frederick M. Hess, *Few States Set World-Class Standards: In fact, most render the notion of proficiency meaningless*, 8 *EDUC. NEXT* 70, 70-72

that states themselves, in a vacuum, want to have high standards and want to test against those standards. This assumption is not unreasonable, as conceivably every citizen in every county of every state *wants* schools to perform well and to teach children effectively. If that assumption is true, then the only reason a state would lower standards—which is really what one is concerned about when discussing the disparity in state standards as no one is likely distraught over the world-class standards set by Massachusetts—is a response to some countervailing disincentive that overwhelms the positive incentive to have high standards.<sup>245</sup>

Hence, the core issue becomes whether NCLB provides disincentives against the creation of higher standards (or *incentives* for lowering standards). The theory is that states would keep their standards lower so that their schools could avoid the consequences of accountability by maintaining a lower bar. It cannot be rationally disputed that there is at least a theoretic material incentive to lower standards under NCLB: states can increase the proportion of their schools meeting AYP if they lower their standards. However, empirically, it is not clear that most states are lowering their standards; rather, comparing all states reveals not a national trend but a hodge-podge of fifty different trends.<sup>246</sup> While Petrilli and Finn argue, forcefully, that states are largely responding to NCLB by lowering standards,<sup>247</sup> the data provided by Peterson and Hess,<sup>248</sup> and more recent data by Peterson and Lastra-Anadón show that the data range from increases in the strength of state standards by as much as 57% for Colorado to decreases of as much as 65.2% for South Carolina.<sup>249</sup> Twenty-three states lowered their standards (by an average of 19.31%) and twenty-five states raised their standards (by an average of 23.08%); the national average was actually an increase in standards by 2.77%.<sup>250</sup>

---

(2008), available at [http://educationnext.org/files/ednext\\_20083\\_70.pdf](http://educationnext.org/files/ednext_20083_70.pdf) (emphasizing the incomparability), with Finn & Petrilli, *supra* note 243, at 49 (emphasizing the disincentive NCLB creates).

<sup>245</sup> Framing the issue this way is useful in that it helps explain why many states—most prominently Massachusetts, but also Missouri, Washington, and Hawaii—have not reduced standards despite having an incentive to do so. Paul E. Peterson & Carlos Xabel Lastra-Anadón, *State Standards Rise in Reading, Fall in Math*, 10 EDUC. NEXT 12, 14 (2010), available at [http://educationnext.org/files/ednext\\_20104\\_12.pdf](http://educationnext.org/files/ednext_20104_12.pdf). It is likely that the disincentive against lowering standards, such as voter pressure, outweighs the material incentive they have to lower their standards. It is also worth noting that this Article does not mean to suggest that states may not want *different* standards for any number of valid reasons. It suggests only that states only have an incentive to lower their standards if some sort of negative incentive outweighs the incentive to maintain high standards.

<sup>246</sup> Peterson & Hess, *supra* note 241, at 71.

<sup>247</sup> Finn & Petrilli, *supra* note 243, at 48-50.

<sup>248</sup> Peterson & Hess, *supra* note 241, at 71.

<sup>249</sup> Peterson & Lastra-Anadón, *supra* note 245, at 13.

<sup>250</sup> Data for two states were unavailable. *Id.* See also KEVIN CAREY, THE EVIDENCE SUGGESTS OTHERWISE—THE PANGLOSS INDEX: HOW STATES GAME THE NO CHILD LEFT BEHIND ACT 3, 5 (2007), available at [http://www.educationsector.org/usr\\_doc/The\\_Pangloss\\_Index.pdf](http://www.educationsector.org/usr_doc/The_Pangloss_Index.pdf) (showing that Alabama “gamed” its standards to hide failures in education policy). Peterson and each of his co-authors calculate differences in performance standards by calculating

As a result, empirically, it is not clear that states are responding consistently to this problem by reducing their performance standards. In fact, standards in reading have generally increased, and, as just shown, standards have increased on average.<sup>251</sup> However, as the country gets closer to the 2014 deadline for NCLB's full proficiency standard, either a dramatic increase in schools failing to meet AYP or a decrease in the standards that define "proficient" appears possible. No evidence yet exists to support that this has happened, but in the future states may believe that if they raise standards, schools will be unable to meet their AYP targets. Therefore, any effective reform should have some mechanism that: (1) decreases the disincentive for higher standards; (2) increases the incentives for higher standards; or (3) both.<sup>252</sup>

One potential solution would be to have something akin to federally-mandated standards so every state would be measured against the same yardstick. However, that option is likely politically unfeasible,<sup>253</sup> as failed past attempts at enacting such mandatory federal standards indicate.<sup>254</sup> It should be noted, however, that while mandatory federal standards appear to be unfeasible, over forty states have now agreed to voluntary Common Core content standards,<sup>255</sup> which may help to alleviate this problem if they also can agree to and implement common performance standards.

Another option would be to place greater emphasis on the NAEP test. Peyser, an opponent of national standards, has suggested that reformers use some coefficient attached to comparative NAEP scores as a factor in the allocation of federal money.<sup>256</sup> This option, though also politically unlikely,

---

the difference between scores on state accountability tests and performance on the NAEP "common stick," which shows over time whether performance standards have increased or decreased relatively. IES performed a similar study and corroborated Peterson's findings. See NAT'L CTR. FOR EDUC. STATISTICS, INST. OF EDUC. SCIS., U.S. DEP'T OF EDUC., MAPPING STATE PROFICIENCY STANDARDS ONTO NAEP SCALES: 2005-2007 (2009), available at <http://nces.ed.gov/nationsreportcard/pdf/studies/2010456.pdf>.

<sup>251</sup> Peterson & Lastra-Anadón, *supra* note 245, at 13.

<sup>252</sup> Even though it is not much of a problem, NCLB reform could not keep the incentive structure the same. If, as this Article suggests, the accountability provisions of NCLB are strengthened, the change will effectively increase the problem of downward pressure on standards since failing to meet AYP would mean more than just public embarrassment. To correct for this, there has to be some countervailing change in the incentive structure to ensure that the incentive to decrease standards (or disincentive for raising them) does not increase.

<sup>253</sup> See Robert Gordon, *Political Realities: To get national standards, leaders need to be bold*, 6 EDUC. NEXT 56 (2006), available at [http://educationnext.org/files/ednext20064\\_56.pdf](http://educationnext.org/files/ednext20064_56.pdf).

<sup>254</sup> Andrew Rudalevige, *The Politics of No Child Left Behind: Lessons from the Clinton years taught Washingtonians that dollars need to be tied to gains in student performance. But did the need to build consensus give too much leeway to state capitols?*, 3 EDUC. NEXT 63, 64 (2003), available at [http://educationnext.org/files/ednext20034\\_62.pdf](http://educationnext.org/files/ednext20034_62.pdf).

<sup>255</sup> Press Release, Nat'l Governors Ass'n, Fifty-One States And Territories Join Common Core State Standards Initiative (Sept. 1, 2009), available at <http://www.nga.org/portal/site/nga/menuitem.6c9a8a9ebc6ae07eee28aca9501010a0/?vgnextoid=1716f7e861ed3210VgnVCM100005e00100aRCRD&vgnnextchannel=759b8f2005361010VgnVCM1000001a01010aRCRD>.

<sup>256</sup> James Peyser, *Hoop Hassles: Incentives, not national control*, 6 EDUC. NEXT 52, 55 (2006), available at [http://educationnext.org/files/ednext20064\\_52.pdf](http://educationnext.org/files/ednext20064_52.pdf).

would certainly reduce the *incentive* to lower standards, since it would decrease the material gains of reducing standards.<sup>257</sup>

A third and perhaps more appealing option to address the disincentives against raising standards would be for federal law to be clearer that states can create a set of tiered standards. Under such a system, states could maintain their current definitions of proficiency and add a new “advanced” level above the current “proficient” level. By setting a more rigorous “advanced” level, states would be able, for example, to implement college- and career-ready standards that are more rigorous than the standards that many states currently have under NCLB, without risking subjection to greater sanctions under NCLB for doing so.

States could then maintain their current standards for the “proficient” level for the time being, which would allow them to continue to strive to bring all subgroups to this basic level of proficiency and would allow for continuity in measuring AYP under NCLB. This would remove the disincentive against raising standards since the higher standard would not be used for AYP purposes. Additionally, states could implement rewards and other consequences that could be tied increasingly to how well schools or districts perform at the “advanced” level, which could incentivize teaching beyond the “proficient” level. Accountability systems could be sensitive as well to achievement at the “advanced” level by subgroups.

For states that raise their standards to “advanced,” perhaps there could be flexibility to extend the 2014 deadline as it relates to “proficient” for a short period of time. There could be a somewhat longer extension for states that have defined “proficient” as on the path to college and career readiness. This transitional approach may ease the problem discussed earlier of over-identification of schools not making AYP under NCLB as 2014 approaches.

### *B. Incentives for Test Exclusion & Grade Retention*

When accountability systems are linked to test results, a major concern is always correcting for the incentive for “test exclusion,” wherein schools find creative ways to prevent poorly performing students from taking part in the testing to artificially boost school test scores. Retaining students on years prior to testing achieves the same effect.<sup>258</sup> This worry is among the more troubling forms of “gaming” testing systems discussed in the literature.

One method of test exclusion made more relevant since the passage of NCLB<sup>259</sup> is placing students into special education programs. Empirically,

---

<sup>257</sup> It would not have an effect against the disincentive for heightening standards, unless, of course, the higher standards themselves induced greater performance on the NAEP.

<sup>258</sup> This is still true when testing is done every year. The poorly performing student goes through the same grade to learn the same material again and would take a test at the same level rather than a more difficult test the following year.

<sup>259</sup> NCLB requires 95% test participation but excludes from that number severely disabled students. 20 U.S.C. § 1412(a)(16)(C) (2006).

the literature is divided as to whether the introduction of an accountability system causes an initial increase in special education placements. According to Jacob, large increases in test scores following the introduction of accountability systems were accompanied by increases in special education placement and by grade retention.<sup>260</sup> Deere and Strayer found the same pattern in Texas.<sup>261</sup> However, Koretz and Barron found no strategic grade retention in Kentucky.<sup>262</sup> Similarly, Carnoy and Loeb found that states that had accountability systems *did not* have significantly higher retention or lower high school completion rates.<sup>263</sup>

In addition to the lack of clear empirical evidence showing systemic increases in retention or test exclusion, there is a very convincing reason why this should not be a source of serious concern: the benefits of test exclusion or grade retention are short-term and do not appear to endure. As Hanushek and Raymond very convincingly show, a school increasing grade retention or test exclusion by placing students in special education creates difficulties for itself in the long term.<sup>264</sup> As they describe the intermediate- and long-term dynamics:

If testing begins in the third grade, the school might exclude this slow student through, say, placement in special education or counseling the student to be absent on the day of testing. If excluded, the average of all remaining students would be higher than otherwise, and the school will tend to look better in comparison to the third grade in the prior year.

But, consider the dynamics. The next year comparison of third grades will be worse because the base comparison has been artificially elevated. Moreover, once excluded, there is a continuing incentive to keep the student out of the testing if subsequent grades are also involved in the accountability system. This continuing incentive puts some restraint into the system, because the school probably cannot increase the exclusion rate year after year. Moreover, since the potential importance of exclusion rates is widely recognized, the school is always at risk that regulatory

---

<sup>260</sup> Brian A. Jacob, *High Stakes in Chicago: Did Chicago's rising test scores reflect genuine academic improvement?*, 3 *EDUC. NEXT* 66, 68-69 (2003), available at [http://education-next.org/files/ednext20031\\_66.pdf](http://education-next.org/files/ednext20031_66.pdf); Brian A. Jacob, *Incentives and Behavior: The Impact of High-Stakes Testing in the Chicago Public Schools* (Nat'l Bureau of Econ. Research, Working Paper No. 8968, 2002), available at <http://www.nber.org/papers/w8968.pdf>.

<sup>261</sup> Eric A. Hanushek & Margaret E. Raymond, *Lessons about the Design of State Accountability Systems* 21-22 (2002) (unpublished manuscript), available at <http://led-pro.stanford.edu/hanushek/admin/pages/files/uploads/accountability.Harvard.publication%20version.pdf>.

<sup>262</sup> DANIEL M. KORETZ & SHEILA BARRON, *THE VALIDITY OF GAINS IN SCORES ON THE KENTUCKY INSTRUCTIONAL RESULTS INFORMATION SYSTEM (KIRIS)* 34-35, 110-11 (1998), available at [http://www.rand.org/pubs/monograph\\_reports/2009/MR1014.pdf](http://www.rand.org/pubs/monograph_reports/2009/MR1014.pdf).

<sup>263</sup> Carnoy & Loeb, *supra* note 38, at 313-22.

<sup>264</sup> Hanushek & Raymond, *supra* note 261, at 19-23.

changes may make it necessary in the future to bring some previously excluded students back into the accountability system.<sup>265</sup>

The point here is that the incentive structure created by accountability systems may already create a built-in long-term disincentive against test exclusion and grade retention to combat the short-term incentives for exclusion and retention. Administrators may be myopic and focus on short-term incentives, but the likely cures for short-term decision-making are probably outside of the scope of anything the federal government could do in federal legislation.<sup>266</sup> Thus, despite its rather prominent place in the literature, the problem of test exclusion and retention has not been shown empirically to clearly exist, nor is it clear that the existing incentive structure needs reform.

There is another issue that is not technically “test exclusion” but has the same effect as test exclusion and *is* a problem under NCLB’s 95%-participation regime. While NCLB requires 95% of the students in each subgroup to participate in testing, this requirement does not apply where the number of members of a particular subgroup (*n*) is insufficient for statistically significant data.<sup>267</sup> As the Thomas B. Fordham Institute shows, states have the discretion to determine the value of *n*.<sup>268</sup> Some states require that a relatively small number of students be in a subgroup in order for that subgroup to be included in NCLB’s accountability mechanisms, and others require a relatively large number. Worse, many states appear to use this discretion to boost the relative number of schools meeting AYP.<sup>269</sup> While this is not test-exclusion as this Article has been defining it, it effectively excludes from accountability the test scores of many minority students.

Unlike the standards dilemma, there is no logical principle for why the subgroup threshold should not be standardized. Developing standards suggests weighing different policy concerns and different values. Further, there is a benefit to allowing the state laboratories to experiment with developing standards. There are no such concerns with determining what constitutes a statistically significant subgroup threshold. If the threshold for statistical significance in State A is 20 black students in a 200-student school, there is no reason why that number should not be statistically significant in a similarly sized school in State B. This is a point of important consideration for future reforms. NCLB’s 95% testing requirement was a great step in preventing myopic test exclusion; future reform could advance this policy by standard-

---

<sup>265</sup> *Id.* at 19-20.

<sup>266</sup> Administrative tenure and the use of longer-scale data by schools to judge administrator pay are examples of possible cures. It is likely both politically unfeasible and inappropriate on the grounds of federalism for the federal government to begin directing schools’ personnel policies in such a broad way as would be required to combat this problem.

<sup>267</sup> 20 U.S.C. § 6311(2)(I)(ii) (2006).

<sup>268</sup> THOMAS B. FORDHAM INST., *THE ACCOUNTABILITY ILLUSION* 17 (2009), available at [http://edexcellence.net/accountability\\_illusion/Introduction%20and%20Methodology.pdf](http://edexcellence.net/accountability_illusion/Introduction%20and%20Methodology.pdf).

<sup>269</sup> *Id.* at 32-46, available at [http://edexcellence.net/accountability\\_illusion/Findings%20and%20Limitations.pdf](http://edexcellence.net/accountability_illusion/Findings%20and%20Limitations.pdf); *id.* at 47-49, available at [http://edexcellence.net/accountability\\_illusion/Discussion.pdf](http://edexcellence.net/accountability_illusion/Discussion.pdf).

izing the subgroup threshold requirements to ensure that “statistical significant threshold requirements” are not used as a proxy for excluding underperforming subgroups from a school’s accountability mechanisms.

### C. Curriculum-Narrowing & “Teaching to the Test”

The third major criticism of accountability systems is that they encourage “teaching to the test” and the narrowing of curriculum to tested subjects at the expense of teaching other skills, such as critical thinking skills or creativity. A 1999 positional paper by the International Reading Association articulates this concern at length, arguing that testing encourages teachers to teach a narrow subset of skills that will increase test performance rather than focus on deeper understanding that can readily be transferred to similar problems.<sup>270</sup> A common counter-argument to the “teaching to the test” critique is, as Jay Mathews succinctly put it:

Those who complain are not really talking about teaching to the state test. Unless teachers sneak into the counseling office and steal a copy, which can get them fired, they don’t know what’s on the test. They are teaching not to the test but to the state standards—a long list of things students are supposed to learn in each subject area, as approved by the state school board.<sup>271</sup>

A criticism different in scope, although qualitatively the same, is that NCLB encourages schools to focus solely on reading and math at the expense of other subjects such as science and music. However, this criticism is rather hollow and should only be a point of reference against reform proposals to the extent that one might be concerned with pre-empting the possibility of a curriculum-narrowing that has not yet actually taken place.

According to Dee and Jacob, NCLB has not hurt science teaching in public schools.<sup>272</sup> Anderson found the same.<sup>273</sup> Hanushek and Raymond do show that scores tend to go up more in subjects that enter into the accountability system than in those that do not,<sup>274</sup> and, in this vein, a report by the U.S. Department of Education shows that, on average, students spent about

<sup>270</sup> INT’L READING ASSOC., HIGH-STAKES ASSESSMENTS IN READING: A POSITION STATEMENT OF THE INTERNATIONAL READING ASSOCIATION (1999), [http://www.reading.org/Libraries/Position\\_Statements\\_and\\_Resolutions/ps1035\\_high\\_stakes.sflb.ashx](http://www.reading.org/Libraries/Position_Statements_and_Resolutions/ps1035_high_stakes.sflb.ashx).

<sup>271</sup> Jay Mathews, *Let’s Teach to the Test*, WASH. POST, Feb. 20, 2006, at A21, available at <http://www.washingtonpost.com/wp-dyn/content/article/2006/02/19/AR2006021900976.html>. Mathews echoes another worthy sentiment: why would teaching to the test or curriculum-narrowing be bad? If the tests measure what students should learn, and the curriculum is tailored to what students should know, what is the problem? *Id.*

<sup>272</sup> Thomas Dee & Brian Jacob, *Evaluating NCLB: Accountability has produced substantial gains in math skills but not in reading*, 10 EDUC. NEXT 54, 59-61 (2010), available at [http://educationnext.org/files/EdNext\\_20103\\_54.pdf](http://educationnext.org/files/EdNext_20103_54.pdf).

<sup>273</sup> Lorin W. Anderson, *Upper Elementary Grades Bear the Brunt of Accountability*, 90 PHI DELTA KAPPAN 413, 414-15 (2009).

<sup>274</sup> Hanushek & Raymond, *supra* note 261, at 22.



twenty minutes more per week in reading and about ten in math in 2006 to 2007 than in 2004 to 2005.<sup>275</sup> However, importantly for this point, *the time devoted to other subjects was virtually unchanged*.<sup>276</sup> In another article written for *Phi Delta Kappan*, Elmore and Fuhrman found that, while narrowing occurred in some states, the exact opposite occurred in others.<sup>277</sup> In Kentucky, for example, their research found that accountability was credited with expanding the content taught to include writing and the humanities.<sup>278</sup> They made similar findings about New York and Vermont.<sup>279</sup>

The point is that, despite the vociferous criticisms by some writers, the “curriculum-narrowing” and the “teaching to the test” problems have not yet revealed themselves in the data. More time is devoted to math and reading, but not at the expense of other subjects. This does not necessarily mean that the issue is irrelevant; increases in time spent on reading and math could eventually begin to result in a decrease in time spent on other subjects. At any rate, states currently control their own testing mechanisms such that if a *particular* mechanism causes narrowing, the reform of that test would necessarily have to be conducted by the state rather than by federal legislation. If NCLB reformers are concerned about the *potential* (although not actual) negative effects on untested subjects, then one possible fix for much of the problem would be to expand testing to include the subjects reformers are worried about falling by the wayside.

#### D. Other Issues

There are a number of other technical points that, while insufficient in themselves to bring about the success or failure of federal education policy, are still relevant to the discussion of proposed reforms. One such technical issue involves the actual workings of NCLB’s school choice and tutoring provisions. As noted in the previous Part, when a school is identified as failing to meet AYP, that school must allow parents to place their students in other nearby public schools or offer supplemental teaching services like

---

<sup>275</sup> JAMES TAYLOR ET AL., U.S. DEP’T OF EDUC., STATE AND LOCAL IMPLEMENTATION OF THE NO CHILD LEFT BEHIND ACT: VOLUME IX – ACCOUNTABILITY UNDER NCLB: FINAL REPORT 139-41 (2010), available at <http://www2.ed.gov/rschstat/eval/disadv/nclb-accountability/nclb-accountability-final.pdf>.

<sup>276</sup> Both of these time periods are after the passage of NCLB but, given the close proximity of NCLB to the first time period, one would have expected curriculum-narrowing—even great curriculum-narrowing—as time goes on and 2014 gets closer. On another note, one might wonder how these increases could take place without decreases in other subjects. The Department of Education report suggests that the increases were accomplished through lengthening the school day, more focused use of existing classroom time, decreases in lunch periods or other non-instructional activities, or that the principal was unaware of reductions in other subjects that teachers made on an individual basis. *Id.*

<sup>277</sup> Richard F. Elmore & Susan H. Fuhrman, *Holding Schools Accountable: Is It Working?*, 83 *PHI DELTA KAPPAN* 67, 71 (2001).

<sup>278</sup> *Id.*

<sup>279</sup> *Id.*

tutoring. This would seemingly create a strong incentive for reform in schools identified as failing, as when students leave, funding follows them. However, Petrilli and Howell both note that a very small percentage of eligible students from failing schools actually make use of these options.<sup>280</sup> Perhaps partly due to a lack of information, only roughly one percent of eligible students took advantage of public school choice options in the 2003-2004 school year and the 2004-2005 school year.<sup>281</sup> Less than twenty percent took advantage of supplemental educational services.<sup>282</sup> As Petrilli explains, informing parents of their options under NCLB is the duty of local school districts, whose interests are at odds with school choice and supplemental teaching services since those districts would have to help bear the costs associated with these policies.<sup>283</sup>

Howell describes several causes for the lack of effective school choice. First, many schools already have high mobility such that schools change composition rather frequently even outside of school choice options.<sup>284</sup> Mobility coupled with the rise in the number of charter schools available in many districts means that students already have “built-in” choice even without the execution of their choice rights under NCLB.<sup>285</sup> Like Petrilli, Howell also describes the “friendly discouragement” from school district officials to parents who would be eligible for school choice vouchers.<sup>286</sup> The lack of efficacy in this policy suggests that the law be changed to require districts to follow the spirit as well as the letter of the law’s intent with respect to parental choice.

Another technical issue involves the AYP scale used by NCLB. The literature describes two poignant criticisms of this scale. First, under NCLB, most states use a static measure of AYP—test scores in a given year—rather than attaching rewards or sanctions to “value added” increases in scores for particular students or particular subsets of students.<sup>287</sup> For instance, a school that adds little value to a brilliant student body is found to meet AYP under NCLB if they meet the state-defined proficiency targets for the year, whereas a school that vastly improves the achievement levels of a disadvan-

---

<sup>280</sup> William G. Howell, *One Child at a Time: An inside look at one city’s efforts to offer families the opportunities promised by No Child Left Behind*, 4 EDUC. NEXT 26 (2004), available at [http://educationnext.org/files/ednext20043\\_26.pdf](http://educationnext.org/files/ednext20043_26.pdf); Michael Petrilli, *Testing the Limits of NCLB: Implementation is not the problem*, 7 EDUC. NEXT 52, 53 (2007), available at [http://educationnext.org/files/ednext\\_20074\\_52.pdf](http://educationnext.org/files/ednext_20074_52.pdf).

<sup>281</sup> Petrilli, *supra* note 280, at 53.

<sup>282</sup> *Id.*

<sup>283</sup> *Id.*

<sup>284</sup> Howell, *supra* note 280, at 28.

<sup>285</sup> *Id.*

<sup>286</sup> *Id.* at 29; *see also* Petrilli, *supra* note 280, at 53 (describing the disincentives for districts to inform parents of their options).

<sup>287</sup> However, the Department of Education did start a pilot program in 2005 that allowed a few select states to integrate growth models into their accountability systems. *See* Press Release, U.S. Sec’y of Educ. Margaret Spellings, Secretary Spellings Announces Growth Model Pilot, Addresses Chief State School Officers’ Annual Policy Forum in Richmond (Nov. 18, 2005), available at <http://www2.ed.gov/news/pressreleases/2005/11/11182005.html>.

tagged student body but does not meet the AYP test score “benchmark” will be found not to meet AYP. In effect, the school with the greater “value added” will be sanctioned while the school with the lesser “value added” will not.<sup>288</sup> For this reason, it may be beneficial for future reform proposals to consider implementing a “value added” approach to accountability.<sup>289</sup> Doing so may also correct for a lack of incentive to “add value” to students whose test scores currently are well above the AYP mark, assuring that they are not neglected in school policy.<sup>290</sup>

A second criticism of the AYP scale is that sanctions are determined based on whether schools fall into one of only two categories: meeting AYP or not. Peterson suggests replacing this with a 4- or 5-point scale (which would roughly correspond to A-F grades) for schools.<sup>291</sup> This proposal should be taken seriously if the complex details in crafting it can be worked out.

## VI. A LOOK TO THE FUTURE: PRESIDENT OBAMA’S BLUEPRINT FOR REFORM AND CHALLENGES POSED BY RECENT CONGRESSIONAL ELECTIONS

This Article began with the history of public education in the United States as it related to the evolution of consequential accountability and then turned to present-day data, characteristics, and criticisms of consequential accountability systems. The natural place to end is with a look to the future. The current administration provided a glimpse at what that future might look like when it proposed, in the wake of its promising Race to the Top program, *A Blueprint for Reform: The Reauthorization of the Elementary and Secondary Education Act*.<sup>292</sup>

---

<sup>288</sup> If the school adds great value to its median student, one has to imagine it is more likely than otherwise to meet AYP, so this theoretical problem should not be over-emphasized.

<sup>289</sup> See Peterson, *supra* note 239, at 51; see also Hoxby, *supra* note 217, at 50 (discussing a change that would take into account the progress students make towards proficiency even if they do not achieve it); Paul E. Peterson & Martin West, *Is Your Child’s School Effective?: Don’t rely on NCLB to tell you*, 6 EDUC. NEXT 76, 78-79 (2006), available at [http://educationnext.org/files/ednext20064\\_76.pdf](http://educationnext.org/files/ednext20064_76.pdf).

<sup>290</sup> See John Cloud, *Are We Failing Our Geniuses?*, TIME (Aug. 16, 2007), <http://www.time.com/time/magazine/article/0,9171,1653653-1,00.html> (last visited Oct. 30, 2010); Peterson, *supra* note 239, at 47. But see Dianne Piché, *Basically a Good Model: NCLB can be fixed*, 7 EDUC. NEXT 57, 58 (2007), available at [http://educationnext.org/files/ednext\\_20074\\_57.pdf](http://educationnext.org/files/ednext_20074_57.pdf); Matthew G. Springer, *Accountability Incentives: Do schools practice educational triage*, 8 EDUC. NEXT 74, 77-79 (2008), available at [http://educationnext.org/files/ednext\\_20081\\_74.pdf](http://educationnext.org/files/ednext_20081_74.pdf).

<sup>291</sup> Peterson, *supra* note 239, at 48. As with the idea of a “growth model,” there are some states that have implemented more nuanced point-scale systems under current law, although they must still either pigeon-hole their system into the “AYP/Not-AYP” NCLB dichotomy or obtain a waiver from the Department of Education to not do so.

<sup>292</sup> OFFICE OF PLANNING, EVALUATION, AND POLICY DEV., U.S. DEP’T OF EDUC., *A BLUEPRINT FOR REFORM* (2010), available at <http://www2.ed.gov/policy/elsec/leg/blueprint/blueprint.pdf>. For an excellent overall summary of the Blueprint, see Chad Aldeman, *Obama’s Blueprint for ESEA Reauthorization*, THE QUICK AND THE ED (Mar. 15, 2010), <http://www>.

The Blueprint proposes many substantive changes to federal education policy enshrined in NCLB. In some ways, it might be easy to overstate the significance of the Blueprint; it is an extremely vague forty-one-page document that neither resembles a statute nor resolves many of the particularities that an eventual statute in this realm will. Nevertheless, the details contained in the Blueprint reveal a number of changes, some at the margins, but some at the very heart of the philosophical underpinnings of current federal education policy.

Much of the balance of this Article will be devoted to an analysis of the administration's Blueprint because it represents the most serious proposal currently on the table for modification of federal education law and policy.

Note must be taken, however, of the recent congressional elections, which produced the potential for dramatic change in the legislative branch's agenda. Republicans have taken control of the House and increased their conservative sway over the Senate. The consequences of this historical shift in power and perspective could be momentous, especially with regard to the use of federal power. Republican campaigns included explicit criticisms of increasing federal involvement in policy areas like health, business regulation, the economy, and, to a lesser extent, education.<sup>293</sup> It is too early to know the ripple effects of the election and beyond the purpose of this Article to speculate about their impact, but they must be noticed.

Whether education policy becomes too hot to handle, gets lost among more pressing issues, becomes a target for reducing federal power, or actually draws attention and action from the Congress simply cannot be predicted. However, as this Article proceeds to analyze the administration's Blueprint, the legislative possibilities remain new and uncharted territory for 2011.

The Blueprint, like all human products, is imperfect, but it should be the continuous goal of education policy analysts to strive for ever-increasingly effective reform, casting a critical eye on reform proposals to judge whether they build on what has worked in the past and correct what has not. In that vein, this Article now turns to analyzing the Blueprint. The Blueprint has a number of positive aspects. However, a very powerful negative, what this Article will describe below as the "5% problem," threatens not only many of the positive aspects of the Blueprint but also many of the steps forward made under NCLB state accountability legislation.

There are several positive elements of the Blueprint, and its encouragement to states to raise their standards can perhaps be given the most praise. This is to be done in two ways: (1) creating a requirement that states have

---

quickanded.com/2010/03/obamas-blueprint-for-esea-reauthorization.html. For a more descriptive but less analytical summary, see EDUCATIONCOUNSEL LLC, THE OBAMA ADMINISTRATION'S "BLUEPRINT" FOR ESEA REAUTHORIZATION: OUTLINE OF KEY POINTS (Mar. 18, 2010), [http://publicpolicyinternational.com/files/Elementary\\_and\\_Secondary\\_Education\\_Acts.pdf](http://publicpolicyinternational.com/files/Elementary_and_Secondary_Education_Acts.pdf).

<sup>293</sup> E.g., Julian E. Zelizer, *Democrats' Best Defense? Good Offense*, CNN (Nov. 15, 2010, 11:38 AM), <http://www.cnn.com/2010/OPINION/11/15/zelizer.democrats.offense/>.

“college- and career-ready” (“CCR”) standards as a condition for the \$14 billion of Title I funding;<sup>294</sup> and (2) changing assessment measures to account for *growth* rather than continuing to primarily use static measures of performance levels.

To adequately highlight the change in the Blueprint’s CCR standards proposal, it is important to remember that there are currently no meaningful federal requirements for state standards—the state must simply provide some set of “challenging” standards, however lax or demanding they may be. The Blueprint changes this by hinging federal funding for states on their choosing either: (1) to develop their standards together with their four-year public university systems with some process of certification from the university system that mastery of the standards ensures that a student will not need to take remedial coursework upon admission; or (2) to work with other states to create rigorous common standards, with material incentives to encourage states to develop common standards.<sup>295</sup>

College and career readiness is an excellent focus for which the current administration deserves praise. However, it should be noted that under NCLB’s current framework, states could implement higher CCR standards *today* on top of the existing accountability framework developed by NCLB. As described in Part V of this Article, states could add more robust “advanced” standards to their accountability systems, while still maintaining current proficiency bars until all subgroups reach that bar.

Perhaps more worrisome, however, the Blueprint appears to contain no requirements regarding the rigor of performance standards. This missing component could be fatal to the very improvements in standards the Blueprint seeks if the higher content standards never “pinch” because of, for example, low performance standards.

Another positive is that the Blueprint keeps NCLB’s requirement that states use testing in their accountability systems. In short, high-stakes testing appears to be here to stay.<sup>296</sup> The Blueprint proposes grants for the development of better state assessments that measure higher-order thinking skills in addition to or in the place of multiple-choice responses.<sup>297</sup> States would get formula-based grants to redesign their assessments in reading and math to ensure that they align with college- and career-ready standards; the Blueprint also would provide grants to help states develop tests in other subjects, such

---

<sup>294</sup> Sam Dillon, *Obama Calls for Major Change in Education Law*, N.Y. TIMES, Mar. 13, 2010, at A1, available at [http://www.nytimes.com/2010/03/14/education/14child.html?\\_r=2](http://www.nytimes.com/2010/03/14/education/14child.html?_r=2).

<sup>295</sup> OFFICE OF PLANNING, EVALUATION, AND POLICY DEV., U.S. DEP’T OF EDUC., *supra* note 292. It is possible, indeed likely, that under such a system, a state could adopt the “standards magnet” approach followed by the successful Texas model. See JERALD, *supra* note 179, at 4.

<sup>296</sup> Rose Garrett, *The New NCLB Blueprint: The Good, the Bad, and the Ugly*, CHANGE.ORG (Mar. 19, 2010), [http://education.change.org/blog/view/the\\_new\\_nclb\\_blueprint\\_the\\_good\\_the\\_bad\\_and\\_the\\_ugly](http://education.change.org/blog/view/the_new_nclb_blueprint_the_good_the_bad_and_the_ugly).

<sup>297</sup> OFFICE OF PLANNING, EVALUATION, AND POLICY DEV., U.S. DEP’T OF EDUC., *supra* note 292.

as science and social studies—though these grants would be awarded on a competitive basis instead of by formula.<sup>298</sup> Assuming that these new assessments are valid, reliable, objective, aligned closely to rigorous content standards, and fit with strong performance standards (all major assumptions), these steps could represent a major advance in testing and in reforming the alleged or theoretical “teaching to the test” or “curriculum-narrowing” criticisms.

Finally, the recommendations for promoting greater teacher effectiveness deserve special commendation. The Blueprint represents a very important breakthrough in both acknowledging how important effective teaching is to improving student achievement and gearing policy regarding teaching toward results. This is a major step forward. It would be an even greater advance if the administration and Congress provided teachers more scientific evidence-based tools, such as the practice guides published by the Institute of Education Sciences,<sup>299</sup> to help teachers be more effective in the classroom.

Now, the negatives: while the Blueprint possesses the above-mentioned significant positive proposals, it also possesses significant negative proposals regarding its *accountability* mechanisms. The negative proposals are in many ways a step backward for federal consequential accountability and, worse, they may dilute many of the positive effects described in the previous sections.

As a whole, the Blueprint drastically weakens accountability under NCLB, which already needed strengthening. One way the Blueprint does this is by removing the school choice option and the supplemental teaching requirement that existed for failing schools under NCLB, which, for reasons already described, have not been frequently exercised by those eligible for them.<sup>300</sup> Failing schools would no longer be required to offer tutoring to eligible students or to offer school choice that would allow parents to remove their children from a failing school in favor of another, successful public school.<sup>301</sup> While the school choice and, to a lesser extent, supplemental services provisions of NCLB went largely unenforced, these provisions provided the potential of options for students in failing schools. Removing these options, rather than increasing their efficacy, is a firm step backwards for the federal consequential accountability system.

Much more troubling is what this Article will refer to as the “5% problem”—under the Blueprint, only the bottom 5% of schools will be subject to *any* sanctions at all. Thus, the vast majority of schools would *not be subject*

---

<sup>298</sup> *Obama offers blueprint for rewriting NCLB*, ESCHOOL NEWS (Mar. 15, 2010), <http://www.eschoolnews.com/2010/03/15/obama-offers-blueprint-for-rewriting-nclb>.

<sup>299</sup> See generally PRACTICE GUIDES, U.S. DEPT. OF EDUC., INST. OF EDUC. SCIS., <http://ies.ed.gov/ncee/wwc/publications/practiceguides/> (last visited Nov. 16, 2010).

<sup>300</sup> See Howell, *supra* note 280, at 28-30; Petrilli, *supra* note 280, at 53.

<sup>301</sup> Dillon, *supra* note 294; OFFICE OF PLANNING, EVALUATION, AND POLICY DEV., U.S. DEPT. OF EDUC., *supra* note 292.

to any consequences whatsoever for failure. States would define their criteria and the nation's schools would be divided into several groups:

[S]ome 10,000 to 15,000 high-performing schools that could receive rewards or recognition; some 10,000 failing or struggling schools requiring varying degrees of vigorous state intervention; about 5,000 schools that would be required to narrow unacceptably wide achievement gaps; and perhaps 70,000 or so schools in the middle that would be encouraged to figure out on their own how to improve.<sup>302</sup>

For the top 5% of schools (as defined by the states), greater flexibility in using federal grants along with material incentives (e.g., for bonuses) are provided. Adding this carrot where NCLB only had sticks is a positive development. In addition, even the sticks in the new proposed accountability measures appear intended to be much more muscular than what was provided for under NCLB. According to current U.S. Secretary of Education Arne Duncan, school districts in the lowest performing 5% of schools have four options in order to qualify for \$4 billion in “turnaround funds”:

- Replace the principal, but keep the teachers and improve the school through professional development, strengthening the instructional program, and extended learning time and other strategies
- Close a school and hire a new principal who can hire back up to half of those teachers
- Close a school and reopen under another new governance
- Close a school and send children to a better school elsewhere<sup>303</sup>

Again, on their face, these options are at least as strong, if not stronger, than those under NCLB.<sup>304</sup> However, this impression is misleading. First, these measures only apply to the bottom 5% of any state's school system. For 85% of schools, there will be no federal accountability measure, neither sticks nor carrots.<sup>305</sup> Furthermore, even for the 5% of schools to which ac-

<sup>302</sup> Dillon, *supra* note 294; see also Greg Toppo, *Education Groups Laud, Criticize Obama 'No Child' Overhaul Plan*, USA TODAY, Mar. 15, 2010, available at [http://www.usatoday.com/news/education/2010-03-14-education\\_N.htm](http://www.usatoday.com/news/education/2010-03-14-education_N.htm).

<sup>303</sup> Rose Garrett, *NCLB Reauthorization: The New Blueprint*, EDUCATION.COM (Mar. 16, 2010), <http://www.education.com/magazine/Article/new-nclb-blueprint>.

<sup>304</sup> Apparently, the New York Times has accepted that these measures are stronger than those under NCLB, for the applicable schools at least. See Dillon, *supra* note 294. The Christian Science Monitor has drawn this conclusion as well. See Gail Russell Chaddock, *Obama's No Child Left Behind revise: a little more flexibility*, CHRISTIAN SCI. MONITOR, Mar. 15, 2010, available at <http://www.csmonitor.com/USA/Politics/2010/0315/Obama-s-No-Child-Left-Behind-revise-a-little-more-flexibility>.

<sup>305</sup> Garrett, *supra* note 296. The authors' hypothesis is that consequential accountability—defined as systems in which there are consequences tied to testing against standards—leads to gains in student achievement. A key aspect of consequential accountability is that the threat of

countability sanctions would apply, the “consequence” can be as minor as replacing a principal. Under NCLB’s current “safety valve” provision, many districts have been replacing principals as their primary method of fulfilling NCLB’s accountability provision.<sup>306</sup> The Blueprint simply would codify several options that are already available under NCLB’s “other” or “safety valve” option.<sup>307</sup>

In effect, a state would be compliant with the law and eligible for all federal funding if it has 5% turnover among its school system’s principals. This is not real accountability, and represents no real step forward from the accountability system created by NCLB. While NCLB may not have been a perfect law, it required that each and every school in all Title I districts across America be measured and face consequences based on whether disadvantaged students made progress toward a standard. What NCLB did, and what should never be changed, was to insist that all schools that receive federal money be held accountable for educating low-income children and minorities. Not 5% of such schools. Not 10% of such schools. But rather *all* schools. Under NCLB, even schools where overall achievement is high have to focus on the children within their schools who have not yet achieved proficiency, bringing greater attention and accountability to these students. The Blueprint would eliminate this focus altogether if those schools are not in the bottom 5% of schools.

The data and literature described at length in this Article show consequential accountability’s virtues such that any reform should *strengthen* accountability, not weaken or dilute it.<sup>308</sup> The current administration is right to focus greater attention on the worst 5% of schools, but to turn off the spotlight on other poorly performing schools with poor students and minorities is a step backwards and a serious mistake. It essentially removes these schools from the pressures of consequential accountability.

It is a well-known principle of law that a right exists only as far as its violation is remedied. Similarly, here, an increase in standards and a more effective set of assessments become meaningless where there are no consequences attached to either. For instance, the emphasis on college- and career-ready standards will mean nothing if there are not consequences for failing to meet those standards for 85% of schools; the only effect is changing the goal, not encouraging and incentivizing the meeting of that goal. Further still, more effective assessments become meaningless when there are not

---

consequences drives schools to higher levels of performance. Without consequences for 85% of schools, the Blueprint, therefore, takes away an essential element of any consequential accountability system by removing the incentives to strive for improvement in schools.

<sup>306</sup> As a reminder, this provides that a failing school may, as an option for accountability, enact “any other major restructuring of the school’s governance designed to produce major reform.” See Mead, *supra* note 51.

<sup>307</sup> *Id.*

<sup>308</sup> This is not to suggest that only sticks should be stronger; more carrots should be encouraged as well, and to the effect that the Blueprint adds carrots where there were only previously sticks, it should be applauded.



consequences attached to the results of those assessments. When combined with the “5% problem,” abandoning current proficiency standards would have the effect of eliminating real “pinching” accountability for the vast majority of students who have not yet achieved proficiency. The cumulative effect would simply be an increase in costs of the federal education policy (in new grants and the described incentive structures) with a decrease in its effectiveness (the lack of consequences neutering the efficacy of standards and assessments).

Yet another problem is that if states move toward CCR standards without maintaining, at least in a transition period, accountability for meeting the lower bar of proficiency, all of the students who have not yet achieved proficiency under the current, more basic standards will be overlooked by the new accountability system. These students, who have not yet cleared the lower bar of proficiency, will simply become part of a larger group of students who have not yet achieved the new, higher standards. In effect, there will be no direct attention to ensure that these students achieve even basic levels of proficiency, thus making growth for these students harder to measure and harder to achieve.

For this reason, the federal accountability system should hold on to the NCLB framework of AYP until the subgroups across the states largely “top out” on proficiency. Adding CCR standards on top of current proficiency levels would maintain accountability for low-performing students, while aspiring to higher standards for all as the new system evolves.

The administration is right to propose taking firmer steps at turning around the lowest-performing schools. This is a good policy that will ensure that the light shines brightest on the school system’s darkest spots. The authors cannot emphasize enough that their criticism of the “5% problem” of the Blueprint is not a disagreement about the administration’s commitment to taking firmer steps with the lowest-performing schools; on the contrary, the authors applaud this and only take issue with the near-complete abandonment of accountability for the rest of public schools.

The lessons of the history of consequential accountability suggest that these aspects of the Blueprint should be changed, lest reformers take two steps back from NCLB’s one step forward.

## VII. CONCLUSION

As this Article shows, consequential accountability systems have grown in both size and influence since their first incarnations in the 1990s. As consequential accountability systems have evolved, policymakers have had the opportunity to examine what works, as well as what areas are in need of improvement. This Article has shown that the data support the general proposition that consequential accountability systems tend to correlate with gains in student achievement. Data from the 1990s are largely positive and indicate that states with consequential accountability systems experienced

statistically significant gains in student achievement relative to states without them. In examining data from the current decade, this Article has shown that the rise in consequential accountability correlates with fairly substantial gains in student achievement. While more research is needed to establish a firm and clear causal relationship, it appears that consequential accountability policies may play an important role in bringing about significant gains in student achievement, especially at the lower grade levels and among the subgroups targeted by accountability systems. Once consequential accountability attaches to high school performance generally across the states, the hypothesis of the positive imprint of consequential accountability can be tested in higher grade levels as well.

As consequential accountability systems have evolved, improvements have been made. However, much work remains to be done. As the country moves into this next era of consequential accountability with the upcoming reauthorization of the ESEA, reformers face a great opportunity to incorporate information learned from both NCLB's successes and weaknesses into the next incarnation of federal policy. Going forward, the current administration's Blueprint and its experiment with the Race to the Top program offer several positive proposals—for example, developing CCR standards, making better use of data, and focusing on fostering teacher effectiveness. While these proposals are excellent steps in the right direction, the Blueprint also gives us great cause for concern—it proposes drastically weakening NCLB's accountability mechanisms to the point where only the bottom 5% of schools would face any sanctions at all for poor performance. As this Article has described, eliminating accountability sanctions for up to 95% of schools will take out virtually the entire “pinch” of consequential accountability, effectively threatening to undermine most, if not all, of the steps forward that the Blueprint proposes.

While NCLB is not a perfect law and a great deal has been learned about how it can be improved as reformers move forward, it seems fairly clear that consequential accountability has the potential to bring about significant gains in student achievement. The authors urge Congress and the Obama administration to strengthen, rather than weaken, federal accountability policies in order to improve the education system until all children achieve true proficiency at the high standards of knowledge and skills required in today's complex and demanding world.