Finding a Common Yardstick: Implementing a National Student Assessment and School Accountability Plan Through State-Federal Collaboration

Marc Pilotin†

INTRODUCTION

Standardized testing permeates American public education. For instance, public school students in California take at least thirty-two state-mandated standardized tests over the course of their K-12 careers.¹ They devote two weeks of the academic year to taking these annual exams,² which can influence their class placements for the following year.³ These two weeks of tests cap a
year of test preparation, in which students may forgo learning about non-tested subjects, such as music and art, and may sacrifice recess to add instructional time to the school day.\(^4\) Students’ scores are shared with their parents and teachers. Eventually, their scores are also aggregated to determine, as required under the federal No Child Left Behind Act (NCLB),\(^5\) whether their schools should be labeled successes or failures.\(^6\)

NCLB has increased the prominence of standardized tests. In order to receive federal education funding, states must administer exams annually.\(^7\) These tests determine whether schools are progressing toward NCLB’s goal of 100 percent student proficiency by the 2013–2014 school year.\(^8\) As that deadline approaches, the pressure on schools increases because those that fail to make adequate yearly progress face sanctions of increasing severity.\(^9\)

These exams, which ostensibly were intended to bolster education quality through accountability,\(^10\) have engendered criticism. Commentators point to the tests’ effect in the classroom. In particular, they note the “narrowing” of the school curriculum: teachers spend substantially more time on tested subjects, to the exclusion of non-tested subjects such as the arts and physical education.\(^11\) Criticisms of the tests’ effects outside the classroom have also been made, which include concerns that the tests have exacerbated interstate disparities in education quality.\(^12\) NCLB allows states to adopt differing definitions of student proficiency.\(^13\) Thus, among the states, students and schools are measured using varying yardsticks, which create unreliable data as to whether students and schools are truly improving. With regard to education quality itself, allowing states to set their own benchmarks creates incentives for states to lower academic standards, particularly because of NCLB’s requirement to sanction schools that consistently fail to raise test scores. Further, the use of...

---

\(^4\) See infra Part I.C.1 (discussing accountability testing’s impact on the school curriculum).
\(^7\) See id. § 6311(b)(2)(A), (b)(3)(A).
\(^8\) See id. § 6311(b)(2)(A); 34 C.F.R. § 200.15(a) (2009).
\(^10\) See, e.g., id. § 6301.
\(^11\) See infra Part I.B.
\(^12\) See, e.g., Goodwin Liu, Interstate Inequality in Educational Opportunity, 81 N.Y.U. L. Rev. 2044, 2091 (2006); infra Part I.C.2.
varying yardsticks demonstrates the arbitrary manner in which schools are labeled: a school labeled failing under one state's standards may be lauded as a success under another state's benchmarks.

With these problems as a backdrop, the Obama administration has embarked on a program to improve student assessment in the United States. As part of the $4.35 billion in education stimulus funding included in the American Recovery and Reinvestment Act of 2009, $350 million has been set aside for consortia of states to develop common assessments. The administration's "Race to the Top" Assessment Program, which will fund the creation of new tests, follows states' efforts to develop common academic content standards. These standards guide what teachers teach and, by extension, what states test each school year.

Because assessments offer insight into whether schools are delivering curricula successfully, it arguably follows that common academic content standards require common assessments. As discussed in further detail below, absent common assessments, states could have overlapping curricula, but evaluating the quality of delivering the standards to students could vary from state to state. In other words, absent an enforcement scheme, there is no assurance that common content standards—on their own—will increase student achievement nationwide. If common standards are not enforced uniformly through a national assessment and accountability system, there is no guarantee that states will perform equally in delivering instruction. Despite common academic standards, educational quality disparities can persist if states remain free to choose what constitutes student proficiency.

Testing programs must be improved, particularly if student assessment begins to play a role in determining teacher pay. Several policymakers have suggested that teacher pay should be tied to increases in student achievement. In particular, President Obama has proposed that Congress increase funding for the Teacher Incentive Fund, a federal program that provides grants to school districts for pilot performance-pay programs. One such program in Denver already relies on student test score growth to help determine teacher pay.

Using student assessment and school accountability reform as its focus, this Comment proposes creating a permanent role for the federal government in coordinating and encouraging education policy innovation. The Race to the Top program provides insight into a structure for interstate collaboration led by Washington; its limited timeframe and scope, however, do not appear to alter significantly the relationship between Washington and the states or the federal government's future role in American education. As outlined below, this Comment asserts that education reform that begins as state experimentation encouraged by the federal government can lead to national education policy that benefits all students. The proposed collaborative relationship between Washington and the states relies on principles of new governance and democratic experimentalism, which leverage the states' traditional roles as democratic laboratories. Such a relationship would proceed in two stages: First, Washington would enable state experimentation in a reform area, such as student assessment and school accountability, and would collect the "best practices" that emerge from the states. Congress can enable experimentation through offering competitive grants, which fund state projects for a limited number of years. The Race to the Top program offers such grants. Second, as best practices become apparent, policymakers would harness states’ best practices to craft national education policy. The goal of this approach is to develop new national education programs that the states have tested. By rooting the components of education reform in state experiments, federal policymakers would collateral benefit by avoiding the political controversy that can accompany innovative assessment reform initiatives.

As noted above, this Comment applies its policy proposal to student assessment and school accountability reform. However, the model suggested here can and should be applied to other areas of education policy. The collaborative model explained below harnesses the strengths of our country's federalist system: the federal government serves as a coordinator and facilitator


22. See New State Ice Co. v. Liebmann, 285 U.S. 262, 311 (1932) (Brandeis, J., dissenting) ("Denial of the right to experiment may be fraught with serious consequences to the Nation. It is one of the happy incidents of the federal system that a single courageous State may, if its citizens choose, serve as a laboratory . . . .").

23. See infra Part III. A citation.


25. Controversy has often arisen in states that have experimented with innovative assessment regimes, particularly those that do not rely on multiple-choice tests. See, e.g., Shelby A. Wolf et al., "That Dog Won't Hunt!": Exemplary School Change Efforts Within the Kentucky Reform, 37 AM. EDUC. RES. J. 349, 353 (2000) (explaining doubts about Kentucky's portfolio assessment system); Jean Merl & Carl Ingram, Wilson's Veto Halts School Testing Program, L.A. TIMES, Sept. 28, 1994, at A1 (reporting on then-Governor Wilson's veto of funding for the California Learning Assessment System, which relied less on multiple-choice exams).
of reform, whereas the states function as the democratic laboratories in which programs are first tested. Once a policy is proven viable at the state level, Washington then works with states to adopt the program nationwide, for the benefit of all students.

Admittedly, because it suggests that state-tested reform should be implemented on a national level, this proposal challenges traditional notions of local control of public education. As the Supreme Court has noted, "[n]o single tradition in public education is more deeply rooted than local control over the operation of schools." However, given the mobility of families and the need for American children to compete with their international counterparts, strict adherence to this tradition can hamper reform.

To be clear, this Comment's primary focus is not the existing NCLB accountability framework. Much can be, and has already been, said about the inadequacies of that law. Although NCLB provides the context for understanding the present state of student assessment, the Comment focuses prospectively on how to develop a new system to evaluate students and schools. In particular, the Comment makes two proposals: (1) a national assessment and accountability plan should be developed and implemented through a cooperative framework between Washington and the states; and (2) this framework, based on democratic experimentalism, should structure future federal-state interactions concerning the development and implementation of education policy.

These proposals are explained in three Parts. Part I briefly reviews the current statutory backdrop, the problems plaguing student assessment, and the Obama administration's Race to the Top Assessment Program. Part II asserts that each level of government must play a role in education and argues for a gradual shift away from local control of education, first by addressing the inequities created by the continued deference to local control and then by establishing the need for coordinated national education reform. Finally, Part III sets out three policy prescriptions rooted in democratic experimentalism that describe how to create a permanent, collaborative relationship between Washington and the states that should be leveraged to develop a new student assessment and school accountability system. In particular, the federal-state relationship (1) would provide states flexibility with and funding for assessment programs in exchange for meeting general federal guidelines, (2)


27. See, e.g., Ryan, supra note 13; Andrew Spitser, School Reconstitution Under No Child Left Behind: Why School Officials Should Think Twice, 54 UCLA L. REV. 1339, 1373–79 (2007) (discussing the problems with using test scores to judge teacher performance); Tamar Lewin, States Found to Vary Widely on Education, N.Y. TIMES, June 8, 2007, at A21 ("Many education experts criticize No Child Left Behind, saying it gives states an incentive to set low standards to avoid sanctions on schools that do not increase the percentage of students demonstrating proficiency each year.")
would improve capacity within the U.S. Department of Education to assist states (or consortia of states)\(^{28}\) in disseminating information regarding their assessment plans, and (3) would establish an independent commission to coordinate national assessment policy.

I

FEDERAL LAW, STUDENT ASSESSMENT, AND ACCOUNTABILITY

A. No Child Left Behind

NCLB aspires to “ensure that all children have a fair, equal, and significant opportunity to obtain a high-quality education.”\(^{29}\) To this end, the Act sets a goal that all K-12 students will meet a state’s definition of proficiency by the 2013–2014 academic year.\(^{30}\) NCLB requires each state to adopt academic standards,\(^{31}\) an accountability system to ensure that schools reach 100 percent student proficiency by the 2013–2014 school year,\(^{32}\) and annual tests that monitor whether its schools are making adequate yearly progress (AYP) toward the full-proficiency goal.\(^{33}\)

Under the Act, schools and school districts\(^{34}\) that do not make AYP for two consecutive years are labeled as in need of “school improvement.”\(^{35}\) If a school or district continues to miss AYP targets, the state is required to take increasingly severe steps toward improving the school’s or district’s performance. For example, the year a school is identified in need of improvement, students must be offered the option to transfer out of the school

\(^{28}\) For example, New Hampshire, Rhode Island, and Vermont developed the New England Common Assessment Program (NECAP), which the three states use to comply with NCLB requirements. See N.H. Dep’t of Educ., About Us, http://www.ed.state.nh.us/education/doe/organization/curriculum/curriculum/NECAP/NECAP.htm (last visited Feb. 2, 2009).


\(^{30}\) 34 C.F.R. § 200.15(b) (2009).


\(^{34}\) NCLB refers to “local educational agencies,” an umbrella term that encompasses all types of public authorities controlling education, including school districts. See 20 U.S.C. §7801. For the purposes of clarity, this Comment refers to such bodies as “school districts.”

\(^{35}\) Id. §6316(b).
or district.\textsuperscript{36} If the school fails to make AYP for five consecutive years, the state must significantly restructure the school or the district, which may include removing the entire staff.\textsuperscript{37}

As noted above, the assessments in mathematics, reading or language arts, and science determine whether a school or district is making AYP.\textsuperscript{38} Although NCLB does not explicitly require these assessments to take the form of standardized tests, the Act's description of acceptable exams leaves little room for any option aside from a standardized test: the law requires that a state's academic assessments "objectively measure academic achievement"\textsuperscript{39} and provide "itemized score analyses."\textsuperscript{40}

**B. Criticisms of NCLB's Overreliance on Standardized Test Scores for Student Assessment**

The need to overhaul assessment and accountability programs arises, in part, from the problems posed by NCLB's reliance on test scores to determine educational effectiveness.\textsuperscript{41} The problem, as the American Psychological Association (APA) states, is that "a test that has been validated only for diagnosing strengths and weaknesses of individual students should not be used to evaluate the educational quality of a school."\textsuperscript{42}

NCLB uses scores alone to evaluate whether a school makes AYP. Conclusions about whether a school is providing "continuous and substantial academic improvement for all students"\textsuperscript{43} is based upon students' performance on exams that assess their "progress toward achieving . . . academic content standards."\textsuperscript{44} Politicians, policymakers, and the media often make conclusions about the quality of schools based on these tests.\textsuperscript{45}

\begin{itemize}
\item \textsuperscript{36} Id. § 6316(b)(1)(E)(i).
\item \textsuperscript{37} Id. § 2316(b)(8).
\item \textsuperscript{38} Id. § 6311(b)(2)(A), (b)(3)(A). At least 95 percent of a school's enrolled students must take the assessment in order for the school to make AYP. Id. § 6311(b)(2)(I)(ii).
\item \textsuperscript{39} Id. § 6311(b)(3)(C)(xiv).
\item \textsuperscript{40} Id. § 6311(b)(3)(C)(xv).
\item \textsuperscript{43} 20 U.S.C. § 6311(b)(2)(C)(iii).
\item \textsuperscript{45} In 2007, former Education Secretary Margaret Spellings commented that a study showing increased test scores across the states "confirm[ed] that No Child Left Behind has struck a chord of success with our nations' schools and students." Sam Dillon, New Study Finds Gains Since No Child Left Behind, N.Y. Times, June 6, 2007, at B7. Upon the release of higher test scores, politicians and policymakers routinely laud schools for their improvements. See, e.g., Editorial, Better Grades, Balt. Sun, July 16, 2008, at 14A; Danielle Gaines, City School Joins State's 800 Club, Merced Sun-Star, Sept. 5, 2008, at A1; Jim Staats, Grand Jury Lauds Improvements at Sausalito Marin City Schools, Marin Indep. J. (San Rafael, CA), July 7, 2008.
\end{itemize}
Uncertainty as to the validity of state assessment data calls these conclusions into question. The well-documented disparities between the states’ improvement on state tests versus their improvement on the National Assessment of Education Progress (NAEP) highlight the danger in correlating test scores with school quality.\(^{46}\) For instance, if conclusions are made about school quality based on the state exam scores, education quality would appear to improve markedly since 2002.\(^{47}\) If viewed through the lens of NAEP results, however, education quality would appear to have only marginally improved over the same time period.\(^{48}\) Of course, even these conclusions assume that improvements in education quality can be equated with rises in test scores. Improvement in test scores can be attributed to a number of sources unrelated to achievement. For example, research shows that students generally improve after repeated exposure to a particular exam format, which underrates conclusions that schools are genuinely making achievement gains.\(^{49}\)

Professor Daniel Koretz outlines three problems that undermine the validity of drawing conclusions about students based on test data. The first is “construct underrepresentation,” which focuses on whether test questions actually measure what they are intended to measure.\(^{50}\) For example, critics in the 1980s panned the use of multiple-choice tests to assess writing proficiency.\(^{51}\) The second problem is “construct-irrelevance variance,” which refers to variations in performance based on irrelevant factors.\(^{52}\) These factors range from questions that require “unrelated background information that some students lack” to administrative conditions that affect students in disparate ways.\(^{53}\) The final problem involves using tests for improper purposes, such as high-stakes accountability, which can lead to teacher efforts to inflate scores, which, in turn, underminds conclusions about actual student performance.\(^{54}\)

Questions of measurement validity and measurement error, on their own, should immediately raise concerns given NCLB’s reliance on test scores to evaluate schools. In addition to these problems, however, are others that challenge the wisdom of NCLB’s single-test assessment scheme,\(^{55}\) including

\(^{46}\) Though states are posting marked improvements on state exams, improvements on the NAEP have been less pronounced. See, e.g., John Cronin et al., The Proficiency Illusion 35 (2007); Nancy Kober et al., Ctr. on Educ. Policy, Has Student Achievement Increased Since 2002?, at 45–46 (2008), available at http://www.cep-dc.org/ (follow “No Child Left Behind” hyperlink; then follow “Student Achievement” hyperlink; then follow “Full Report” hyperlink within “Has Student Achievement Increased Since 2002?” section).

\(^{47}\) See Kober et al., supra note 46, at 25.

\(^{48}\) Id. at 45–46.

\(^{49}\) See Linn, supra note 41, at 7.

\(^{50}\) Koretz, supra note 41, at 220.

\(^{51}\) Id.

\(^{52}\) Id. at 221.

\(^{53}\) Id.

\(^{54}\) Id. at 220, 235–59.

\(^{55}\) Koretz, supra note 41, at 147.
potential test bias disadvantaging particular subgroups. Moreover, these validity issues may provide legal footholds for attacking performance pay programs for teachers. For all these reasons, future education reforms must include rethinking our assessment process.

C. Increasing Inequality Under NCLB’s School Accountability Provisions

In addition to the problems inherent to the use of standardized tests for student assessment, academics and media commentators alike have identified numerous problems with school accountability under NCLB. These critiques span various subject areas, many of which extend far beyond the scope of this Comment. However, commentators point out that NCLB creates incentives that exacerbate existing inequality among the states concerning education quality.

Broadly speaking, these criticisms fall into two categories: the first concerns testing’s effect within the classroom, particularly on how high-stakes testing in specific subjects encourages schools to narrow their curricula; the second concerns testing’s effect outside the classroom, especially how the states’ ability to define “proficiency” distorts inter-school comparisons.

I. Internal Effects: Test Preparation Narrows the School Curriculum

Annual standardized testing for accountability purposes forces schools to focus on tested subjects, leaving social studies, the arts, and physical education
out in the cold. Quite understandably, if the possibility of state intervention depends on test scores, rational administrators will choose to focus on the tested subjects.\textsuperscript{60}

A recent study shows that, since NCLB's enactment in 2002, 62 percent of school districts have increased their instructional time for language arts and/or mathematics in their elementary schools.\textsuperscript{61} Districts, however, have not increased the length of the school day. Thus, to accommodate the increase in instructional time for reading and math, districts reduce the amount of time spent teaching social studies, science, arts and music.\textsuperscript{62} One-fifth of schools cut students' recess times by an average of fifty minutes per week.\textsuperscript{63}

Reports suggest that students in low-performing schools, which often enroll economically disadvantaged students, most often face the simplified two-subject curriculum.\textsuperscript{64} In other words, educational and extracurricular opportunities are not equal between schools based not only on financial resources, but also on test scores.\textsuperscript{65} In defense of the "meat-and-potatoes" curriculum, some argue that students need a stripped-down curriculum so that they may focus on their basic math and reading skills.\textsuperscript{66} Moreover, so the argument goes, a basic math-reading curriculum enables schools to hold out electives courses as incentives for students who do well.\textsuperscript{67} While teaching such a pared down curriculum may have these positive effects, it nevertheless

\textsuperscript{60} ROBERT L. LINN, CTR. FOR THE STUDY OF EVALUATION, NAT'L CTR. FOR RESEARCH ON EDUC., STANDARDS & STUDENT TESTING, CSE REPORT No. 601, ACCOUNTABILITY: RESPONSIBILITY AND REASONABLE EXPECTATIONS 4 (2003), available at http://www.cse.ucla.edu/products/Reports/R601.pdf (discussing impact of high-stakes testing on curriculum narrowing).


\textsuperscript{62} MCMURRER, supra note 61, at 7 tbl.4.

\textsuperscript{63} Id.


\textsuperscript{65} See William S. Koski & Rob Reich, When "Adequate" Isn't: The Retreat from Equity in Educational Law and Policy and Why It Matters, 56 EMORY L.J. 545, 587 (2006) ("[A] growing body of evidence suggests that new accountability regimes have the effect of narrowing the curriculum and reducing the educational opportunities of students in poor-performing schools, i.e., those with poor and minority students."); Erik W. Robelen, States Face Budget Hurdles on Testing, GAO Says, EDUC. WEEK, Oct. 14, 2009, at 19 (quoting John Tanner, an official with the Council of Chief State School Officers, who stated that "[o]ne of the consequences of test-based accountability is that, particularly in low-performing schools, the test becomes the curriculum, because passing that test becomes so all-important").

\textsuperscript{66} Dillon, supra note 64, at 1.

\textsuperscript{67} Id.; see also LINN, supra note 60, at 4.
exacerbates the existing gap in education quality between high- and low-performing schools: schools that are not subject to NCLB-mandated state intervention may be freer to teach enriching subjects, such as arts and music.

Congress has tried, albeit unsuccessfully, to remedy this problem. During the attempt to reauthorize NCLB in 2007, Congressmen George Miller and Buck McKeon proposed five-year grants for school districts to increase instruction in non-tested subjects, including music, the arts, foreign languages, and physical education. While this proposal provides districts with the opportunity to enlarge their curricula, the opportunity may be illusory: districts would still be subject to the annual testing requirements that narrow the curriculum in the first place. Unless assessment and accountability systems are reformed at the state or federal level, local districts will remain motivated to focus on tested subject areas.

2. External Effects: Disparities in Education Quality Among States

NCLB does not define what constitutes “proficiency”; instead, it leaves that determination to the states. Nor does the law define how much progress schools must make on an annual basis toward the goal of 100 percent proficiency. Without a standard definition of proficiency, states have “perverse incentives” to tinker with their definitions of proficiency. Thus, depending on their circumstances, a state may lower its proficiency requirements, which results in inconsistent standards across the states and thereby worsens existing inequities in educational quality.
States have at least two incentives to minimize what a student must know in order to be labeled proficient. The first incentive involves the sanctions under NCLB itself. As noted above, a school that does not meet annual yearly progress toward the goal of 100 percent proficiency can be subject to federally mandated sanctions. To avoid these sanctions, a state could lower the bar for students so that its schools are not labeled failures. Immediately following NCLB's enactment, states responded by revising their definitions of proficiency to ensure that more of their students were labeled proficient for the purposes of NCLB. A recent study confirms that this trend continues.

States also have a public-relations incentive to ensure that students can easily meet the proficiency standard. A higher standard, which fewer students would meet, would give the public the impression that students are not doing well. Such impressions would not bode well for the state, particularly with regard to its ability to attract businesses, or to politicians, who often tout advances in test scores as evidence of improvements in the quality of a public school system.

The tension between setting high standards and these perverse incentives was on display in California's recent efforts to revise its algebra requirements. The State Board of Education—at the behest of Governor Arnold Schwarzenegger—voted to require that all eighth graders take Algebra I. The vote immediately raised the ire of State Superintendent of Public Instruction Jack O'Connell, who pointed to the problems with getting students to proficiency in Algebra I in eighth grade without additional resources. In a strongly worded press release, O'Connell stated that the more rigorous requirement "set[s] our schools up for failure," pointing to the existing 23 percent proficiency rate in mathematics for seventh graders. Though not explicit, O'Connell's criticisms evince a concern that more schools would be

---

73. See Ryan, supra note 13, at 947-48.
74. See id.
76. See Ryan, supra note 13, at 949 (discussing the consequences offsetting ambitious goals that students cannot meet).
77. Id. at 950.
80. Linda Jacobson, California Board Mandates Algebra 1 for All 8th Graders, EDUC. WEEK, July 16, 2008, at 19.
82. Id.
labeled failures if eighth graders were required to take and be proficient in Algebra I.

With its inconsistent definitions of proficiency, the current assessment and accountability system does not encourage a “race to the top” with regard to educational achievement. On the contrary, there are incentives to keep proficiency requirements low. To address this problem, along with problems caused by efforts to meet testing standards, a new assessment and accountability paradigm is needed, which can be achieved through the proposal discussed below. However, developing a nationwide definition of proficiency requires countering existing traditions involving local control of schools.

II
INEQUALITY AND THE NEED FOR FEDERAL ACTION

A. The Tradition of Local Control over Education

The United States has a strong tradition of maintaining local control over schools. The tradition likely arose from the local funding of education, which inevitably led to a desire for communities to retain control over schools. As the Supreme Court opined, “local autonomy has long been thought essential both to the maintenance of community concern and support for public schools and to quality of the educational progress.” Notions of local control are explicitly codified in federal law: federal control over local curricular matters is expressly prohibited.

NCLB attempts to strike a balance between creating an increased role for Washington and preserving local control. As noted above, the law leaves states in charge of determining what they will teach, what will constitute proficiency, and which tests will be used. However, as already discussed, this exacerbates inequalities among the states.

B. Challenging the Local Control Paradigm

Continued adherence to local control of education allows, among other things, a student to be proficient under one state's standards while not proficient under those of another state. Though researchers and policymakers lament that the nation—as a whole—is at risk, a state-by-state approach to education is tolerated, in part, in the name of local control.

85. See, e.g., 20 U.S.C. § 6575 (2006) (“Nothing in this subchapter shall be construed to authorize an officer or employee of the Federal Government to mandate, direct, or control a State, local educational agency, or school’s specific instructional content, academic achievement standards and assessments, curriculum, or program of instruction.”). This prohibition is repeated at least nine times throughout the United States Code.
A national assessment plan with a uniform proficiency definition can ameliorate these disparities by setting de facto national standards for what students must achieve. Although Washington would not define what states need to teach, using the same yardstick to measure educational outcomes can affect how schools instruct their students. States would be required to meet a national proficiency standard and could adjust their curricula accordingly.

Proposals for national testing in education are not new. Both the George W. Bush and Clinton administrations proposed voluntary national standards and examinations. In the 1990s, the Goals 2000 Act came close to establishing a national test. The Act created the National Education Standards and Improvement Council (NESIC), which had the purpose of certifying and reviewing national content and student performance standards. After Republicans took control of Congress in 1994, however, NESIC and plans for a national test were scrapped. While President Clinton discussed national standards during his last term, the concept died quietly.

Although previous failed attempts to implement national standards could weigh against trying to implement a national assessment and accountability plan, there are indications that times have changed. The National Governors Association, American Federation of Teachers, and various other education groups now endorse the idea of common expectations. Moreover, the 2007 congressional proposals during the NCLB reauthorization effort suggest acceptance of an expanded federal role in education. None of the legislative proposals called for a complete repeal of the law. Instead, many of the bills called for giving states flexibility in how they calculate AYP, which would allow schools to avoid sanctions. Indeed, two bills called for establishing voluntary content standards in mathematics and science and for assessment in those fields. However, Congress did not substantially act on these proposals.

88. Id. § 211(1).
90. See MCANDREWS, supra note 86, at 160–61.
91. See David J. Hoff, National Standards Gain Steam, EDUC. WEEK, Mar. 4, 2009, at 1.
92. See, e.g., H.R. 2946, 110th Cong. (2007) (allowing schools to exclude some students with limited English proficiency from the AYP calculation); H.R. 2577, 110th Cong. (2007) (enabling states to use alternative models for measuring AYP); H.R. 2166, 110th Cong. (2007) (allowing states to exclude the results of students with limited English proficiency); H.R. 2087, 110th Cong. (2007) (enabling states to adopt alternate ways of calculating AYP); H.R. 1711, 110th Cong. (2007) (allowing tracking of AYP on a longitudinal basis); H.R. 648, 110th Cong. (2007) (lowering the percentage of students who need to improve for a school to avoid sanctions).
and took no action on reauthorization during the 2008 election year.\textsuperscript{94}

Further, polls show that Americans support a national education plan. The 2008 Phi Delta Kappa/Gallup Poll found that half of respondents supported national standards.\textsuperscript{95} This result reflects a slight increase over the number supporting national standards in 1994, when President Clinton began advocating for his education agenda.\textsuperscript{96} A study by the Hoover Institution showed even stronger support: 73 percent of respondents to a 2007 survey supported national education standards and a national test.\textsuperscript{97}

Public support and congressional acquiescence to the new federal role suggest that a national assessment accountability plan would fare better today.

\textbf{C. Reasons for a National Assessment and Accountability Plan}

The benefits of establishing a national assessment and accountability plan are numerous. First, a national plan can lower states' costs of assessment. Second, a national plan can leave states with sufficient flexibility to determine their own curricula, thereby preserving local control.

\textit{1. A National Assessment System Can Reduce Development and Administrative Costs for Tests}

A national assessment system can reduce state spending on testing. A national testing regime would eliminate the need for fifty states to develop fifty independent different testing programs and would reduce the associated costs.\textsuperscript{98}

Cost estimates for states on standardized testing run from $517 to $750 million per year.\textsuperscript{99} Per student, the cost is about $10 to $30.\textsuperscript{100} This amount does not include the money schools spend on test-preparation materials.\textsuperscript{101} Tests mandated by NCLB comprise one-third of the standardized testing industry, and, in the years since NCLB's enactment, the industry has experienced substantial annual growth.\textsuperscript{102}

\textsuperscript{94} See Dillon, supra note 71.

\textsuperscript{95} William J. Bushaw & Alec M. Gallup, Americans Speak Out — Are Educators and Policy Makers Listening?, PHI DELTA KAPPAN, Sept. 2008, at 9, 13–14. Researchers found that if the phrase “common expectations” were used in lieu of “national standards,” the percentage supporting the idea jumped to 62 percent. Id.

\textsuperscript{96} Stanley M. Elam et al., The 26th Annual Phi Delta Kappa/Gallup Poll of the Public’s Attitudes Toward the Public Schools, PHI DELTA KAPPAN, Sept. 1994, at 41, 48.

\textsuperscript{97} William G. Howell et al., What Americans Think About Their Schools, EDUC. NEXT, Sept. 22, 2007, at 12, 15.


\textsuperscript{100} Id.

\textsuperscript{101} Id.

\textsuperscript{102} Nightly Business Report (PBS television broadcast Feb. 18, 2008) (transcript on file
The high cost of developing exams has led states to use multiple-choice questions, which are cheaper to develop.\textsuperscript{103} States admit that the use of cheaper multiple-choice questions prevents them from assessing "highly cognitively complex material," which has led some to drop challenging academic content standards because they cannot be tested or to include such standards for "instructional purposes only."\textsuperscript{104}

By creating a common system, states can share costs of the development and administration of assessment systems. Instead of each state bearing the expense of annually updating and revising tests, the country can share these fixed costs. Take, for example, the New England Common Assessment Program (NECAP) developed and administered by New Hampshire, Rhode Island, and Vermont. By collaborating, each state saved about one-third of the cost that would have been spent had they not collaborated.\textsuperscript{105} As a result, NECAP tests contain a combination of multiple-choice and open-response questions; these states’ ability to include open-response questions can be attributed, in part, to the lower costs attained through interstate collaboration.\textsuperscript{106} Furthermore, because it would contract for a large number of tests, Washington could negotiate a lower per-unit cost for the exams. Finally, if the various states administer the same exams, they could share best practices with one another, resulting in greater efficiency.

2. States Will Retain Flexibility to Determine How to Achieve Standards

In creating a national assessment and accountability system, states’ existing content standards would be left largely intact. These new assessment systems, which would require some form of national testing, would necessarily entail setting performance-based standards to ensure accountability. Performance standards detail what students need to be able to do to show proficiency in a given domain.\textsuperscript{107} Unlike content standards, performance standards do not describe "the knowledge, skills, and other understandings that schools should teach."\textsuperscript{108} Stated plainly, performance standards define where students must be at the end of a grade level, whereas content standards explain

\begin{footnotesize}
\begin{enumerate}
\item \textsuperscript{104} \textit{Id.} at 20–21.
\item \textsuperscript{105} \textit{Toch, supra} note 99, at 21. The Race to the Top program embraces state collaboration in assessment development. As discussed further below, the grant program does not accept grant applications for assessment development from individual states. Instead, only groups of states, like the group in NECAP, may apply for grant funding. \textit{See infra} notes 180–84 and accompanying text.
\item \textsuperscript{106} \textit{Gen. Accounting Office, supra} note 103, at 24.
\item \textsuperscript{107} Karen K. Wixson et al., \textit{The Challenge of Developing Content Standards}, 27 Rev. Res. Educ. 69, 73 (2003).
\item \textsuperscript{108} \textit{Id.}
\end{enumerate}
\end{footnotesize}
how students are to get there during the course of the year.

The performance standards involved in a national assessment system would set overall expectations that students would need to meet at various grade levels. By defining performance standards as opposed to content standards, the federal government would only set broad expectations; states would be left to fill in the blanks and determine how best to achieve the federal goals. The federal government has taken a similar approach in the environmental context with the Clean Air Act. Under the Clean Air Act, the federal government sets National Ambient Air Quality Standards (NAAQS) for various air pollutants. States are left to determine how they will meet these standards. Similarly, the federal government can set performance standards for students—thereby creating a common yardstick by which all states would be judged—leaving the states to determine how to shape their curricula in order to get students to meet those standards.

An alternative to performance standards would be national content standards, which define what students must learn in a given grade level. There are two primary reasons why content standards would prove politically unpalatable. First, creating content standards would require the federal government to define exactly what must be taught in every grade level, which explicitly contravenes various federal statutes prohibiting Washington from controlling the curriculum. Second, and perhaps more practically, states have already invested time and money in developing their own content standards.

Of course, some may deride performance standards and a national test as setting a de facto national curriculum: because students would need to perform at a certain level on the exam, states would align their content standards with the needs of the test. This alignment is to be expected. National performance standards leave states to determine their curricular details, but some uniformity among what states teach would be an expected—and, indeed, desired—result.

Finally, some would object to a national proficiency standard on the basis that developing a consensus-based standard would set a low bar for student performance. This need not be the case. The National Council of Teachers of Mathematics developed suggested mathematics standards in 2000, which are

110. See id. § 7409.
highly regarded. Further, today, unlike in 1994, existing state standards can serve as exemplars for national standards. As discussed more below, a collaborative process under the watchful eye of business and other peer states can prevent a regression in student expectations.

III

A Collaborative Approach to Developing New Assessment and Accountability Systems

To develop new assessment and accountability systems, Washington should harness the power of the states as democratic laboratories. In doing so, the goal is to find the "best practices" in developing assessment and accountability systems and in disseminating the information among the states so that they may implement these practices. The best practices identified through this framework would be used to begin implementing a national assessment and accountability plan.

As an initial matter, this Comment uses the phrase "best practices" to refer to discrete techniques that, if implemented, would yield better results than others. As one commentator describes them, best practices are "those actions that surpass all others in pursuit of an agreed-upon goal . . . according to some objectively measurable standard." In the context of the problem of assessment and accountability, these would be the actions that result in positive educational outcomes. Given the scope of this Comment, this definition is necessarily vague; the education experts who would be involved with the process would be charged with determining which are the best practices.

In order to coordinate state efforts in developing plans, this Comment proposes three policies: (1) expanding federal support of research and development in assessment, (2) continuing to build capacity within the U.S. Department of Education to disseminate information about progress being made by states to their peer states, and (3) establishing an independent board of experts that would promulgate regulations based upon the best practices that emerge from the state projects. These policy suggestions arise out of democratic experimentalism theory.

117. See id. (stating that those in the relevant industry decide what constitutes "best practices").
A. Structured Collaboration Through Democratic Experimentalism

Democratic experimentalism falls into the ever-growing body of scholarship on "new governance." New governance provides an alternative to the traditional command-and-control paradigm,119 which is generally understood to involve government-set standards, developed by a legislative or an administrative body, followed by private actors.120 Proponents of new governance would employ a more interactive relationship between regulators and regulated parties, which would lead to more effective regulatory policy.121 The relationship would be "presumptively decentralized," with states and localities taking the lead in developing policy solutions.122 As described by one commentator, this system provides "centrally coordinated local problem solving."123 In other words, democratic experimentalism holds the potential to effect widespread change while respecting notions of local control; in the context of education reform, the theory offers a balance between the need to coordinate national reform and the need to respect local control.

Democratic experimentalism redefines the structure of the relationship between the national government and the states. Experimentalism achieves policy goals through "decentralization, deliberation, deregulation, and experimentation."124 Because an ideal (or even near ideal) assessment and accountability system has not yet emerged, I propose an explicit national policy of research and development through experimentalism.

1. Components of Democratic Experimentalism

The democratic experimentalist model provides a viable framework for this collaborative relationship. This model requires four primary components: setting national goals, benchmarking, simultaneous engineering, and error correction through information sharing.125 Under the first component, Congress sets national priorities and provides funding to the states to innovate and create solutions.126 For instance, with respect to assessment and accountability reform, Congress could require states to, among other things, create systems that alleviate curriculum narrowing and allow for interstate comparisons.127 These

120. Id. at 340.
121. See id. at 822; see also Jody Freeman, Collaborative Governance in the Administrative State, 45 UCLA L. REV. 1 (1997) (arguing for having parties share responsibilities throughout the rulemaking process).
122. See Dorf & Sabel, supra note 21, at 340.
125. See Dorf & Sabel, supra note 21, at 287.
126. Id. at 341.
127. See supra Parts I.B and I.C.
priorities should be sufficiently specific, so as to constitute a discernable goal, but nonetheless broad enough to give states discretion in adopting innovative assessment techniques and even encourage such experimentation.

Benchmarking, the second component, involves states comparing themselves to each other to determine relative success. This process is intended to reveal disparities among the outputs of systems in order to encourage improvement, and can occur at the beginning of the experimentation process—to set a baseline—and at regular reporting intervals. This process requires coordination at the federal level through an administrative agency, such as the U.S. Department of Education. The Department, in conjunction with the states, would develop metrics to be used in evaluating assessment and accountability programs. The Department would then use these metrics to evaluate the programs and would publicly disseminate the results. Public dissemination of information would serve two purposes: First, benchmarking data gives states information that they can use to begin their improvement process. By being able to compare results, the states would “learn by monitoring” and would select best practices to incorporate into their systems. Second, benchmarking provides information to the general public, who can then apply pressure upon policymakers to enact improvements into their state’s system. Indeed, both of these purposes are borne out by empirical research, which shows that time-constrained officials are more apt to consider visible and politically salient innovations.

Simultaneous engineering and error correction, the third and fourth components, are intertwined. Simultaneous engineering means that states would concurrently develop new assessment schemes. As they begin to do so, the states would continue to provide feedback to the Department of Education on their progress. This information, after being compiled by the Department, would be disseminated back to the states to use in the process of

129. Dorf & Sable, supra note 21, at 345.
130. Id. at 347.
131. Id.
132. The Race to the Top program provides some insight into how these metrics could be developed and could function. For that program, the Department of Education developed scoring criteria through notice-and-comment rulemaking to judge grant applications. See Race to the Top Fund Final Rule, 74 Fed. Reg. 59,688, 59,808 (Nov. 18, 2009) (to be codified at 34 C.F.R. pt. B, ch. 2). Scorers, taken from the education community, will be recruited to judge applications. See Race to the Top Fund Overview Information and Notice Inviting Applications for New Awards, 74 Fed. Reg. 59,836, 59,845–46 (Nov. 18, 2009).
133. See Dorf & Sable, supra note 21, at 287 (emphasis omitted).
134. See Andrew Karch, Democratic Laboratories 9 (2007).
135. See id.
error correction.\textsuperscript{136} Stated plainly, error correction involves states learning from the mistakes of their peers. Similarly, states would refine their own systems by incorporating other states' best practices.\textsuperscript{137} For example, if California finds that an assessment system incorporating essay questions results in inconsistent grading, Iowa can learn from California's experience and take steps accordingly. The experience of one state would become shared knowledge for all.

Information provided by professional associations and think tanks can also augment the body of knowledge used by policymakers. Empirical evidence shows that these types of organizations, such as the National Governors Association and Achieve, provide lawmakers with better information.\textsuperscript{138} To this end, information provided to the Department of Education must also be made available to these nongovernmental organizations that can also process the information and provide feedback to the states.

This Comment's proposal limits its use of democratic experimentalism to these four points. Admittedly, this proposal conflicts with some precepts of experimentalism, particularly to the extent that it relies on the theory for the limited goal of creating discrete policies, which, for the purposes of this Comment, involve student assessment and school accountability. The experimentation here is intended to yield a single solution to be adopted and implemented nationally. Generally, experimentalism envisions an enduring decentralized system for all purposes.\textsuperscript{139}

2. Criticisms of the Experimentalist Paradigm

Democratic experimentalism is not without its critics. One kind of criticism, endemic to any type of purposeful incremental reform, is that the process would plod too slowly and would thereby delay the achievement of policy goals.\textsuperscript{140} In the context of assessment reform, however, the potential for delay should be tolerated, because the process is focused on researching and developing the best assessment program possible. Unlike other policy contexts where the process primarily focuses on implementation, the experimentalist project in assessment systems involves both development and implementation. The experimentalist project is iterative. As policymakers execute their assessment plans annually, we should expect that they come back the following year with changes that reflect what they have learned.

Congress has at least two options to ensure that some progress is being made, although these options admittedly contravene the pure experimentalist paradigm. First, Congress could set deadlines, which will inherently limit the experiment. Second, Congress could mandate that the states begin adopting the

\textsuperscript{136} See id.
\textsuperscript{137} See Sabel, supra note 118, at 123.
\textsuperscript{138} See KARCH, supra note 134, at 4–9.
\textsuperscript{139} See Dorf & Sabel, supra note 21, at 340.
\textsuperscript{140} See id. at 404–05.
best practices that emerge. This second option, for which I advocate below, also constrains the experimental endeavor by slowly ratcheting up the command-and-control regulation that Congress would impose on the states. A pure experimentalist would likely object to such an approach. However, because this proposal aims at creating a uniform assessment program, it necessarily departs from pure experimentalism insofar as the theory imagines ongoing state control of the project.

Professor David Super critiqued experimentalism in his article about the failure of antipoverty law. Because his critique highlights the pitfalls of experimentalism, I rely on his list of criticisms as a guide for my policy proposals. Professor Super highlighted six assumptions that should be addressed in any experimentalist project: (1) that there is a general consensus about the existence and nature of the problem, (2) that all relevant players are “inclined to act in a public-spirited way to correct that problem,” (3) “that reliable metrics exist,” (4) “that time does not constrain decision making,” (5) that national regulation is not needed, and (6) that “transaction costs do not significantly deter political participation.” If any one of these assumptions proves to be false, Professor Super argues, then the particular experimentalist project will fail.

The third assumption concerning metrics seems to pose the most substantial challenge to this proposal. Benchmarking assessment systems requires a valid measure that can be used to evaluate the assessment programs. How would states compare their systems? Would a system’s success be based upon student feedback? Should student performance on the assessment play a role? Would teachers’ opinions be taken into account? To be sure, any set of metrics must account for the problems that motivate the assessment experiment in the first place (i.e., over-testing, curricular narrowing, and a failure to align with workplace expectations and international benchmarks). Additionally, to combat the likelihood that one state’s definition of a successful system will not align with that of other states, the benchmarking metrics should be set by Washington in consultation with the various states.

Professor Super’s concern that all players will act in a “publicly-spirited way” must also be addressed. He argues that experimentalists assume that opportunists and “recalcitrant perpetrators,” who seek to oppose any action, will not be effective in a democratic experimentalist project. He doubts that the efforts at transparency would be sufficient to prevent capture of the process by special interests. In the assessment context, many players could fit this mold. Teachers’ unions, which have often been singled out for challenging

141. See, e.g., Super, supra note 124.
142. Id. at 553–58.
143. Id. at 559.
144. Id. at 554–55.
145. Id. at 555.
reform efforts, reform efforts, could be portrayed as the "recalcitrant perpetrators," and the testing industry as the opportunists, seeking to gain a market for their own products. Apart from these two groups, other stakeholders—including parents, administrator associations, and school board associations—could create roadblocks. Thus, Professor Super’s criticism applies in this context, although it is difficult to imagine any governance structure impervious to interest group politics. I address his concern below, where I propose creating an independent board overseeing this project that includes stakeholders from various groups. Instead of avoiding the influence of interest groups, I propose that their representatives be explicitly brought to the table.

The other assumptions highlighted by Professor Super are less problematic. Two assumptions—the need for unlimited time and the lack of a need for national regulation—do not apply. As mentioned above, this project is not meant to be an ongoing, deregulatory endeavor. Instead of assuming unlimited time and deregulation, my proposal diverges from the experimentalist paradigm by explicitly setting a deadline and incrementally setting national regulation.

Because public participation is not central to my proposal, Professor Super’s concern regarding the transaction costs of such participation is inapposite. While there will be provisions for public input, as in any administrative policymaking, public involvement is not an integral part of my proposal.

Finally, as to Professor Super’s concern about whether there is a general consensus concerning the problem, I believe that there is. Policymakers and stakeholders, from Education Secretary Duncan to the teachers’ unions, believe that the current assessment system is insufficient. Although there may be disagreement as to what the optimal system would be, there is agreement on the need to work toward finding that system.

B. Proposal for a Collaborative Approach

I. Financial and Regulatory Incentives for States to Develop and Adopt New Assessment and Accountability Programs

To implement the democratic experimentalist proposal outlined in this Comment, Congress must authorize and fund the democratic experimentalist assessment project. There are three main components to consider in drafting

146. See, e.g., David Leonhardt, The Big Fix, N.Y. TIMES, Feb. 1, 2009, at MM22 (naming the teachers’ unions as part of a “murderers’ row of interest groups”).
this authorization: the breadth of the authorization, incentives to participate in
the program in light of current NCLB requirements, and the length of
the project.

To encourage state participation in this project, Congress could couple the
examination requirement with a release from NCLB’s current requirement that
students be assessed on an annual basis for the purposes of accountability. That
is, states could continue to assess on an annual basis—but to diagnose student
needs, not for accountability purposes. Such an approach would address the
problem of curricular narrowing that results from annual high-
stakes examinations.

If Congress lifted the annual testing requirement for accountability
purposes, however, the NCLB accountability model must change because
yearly progress could no longer be measured. This shift in testing focus may
not be fatal. Because states would have flexibility in how to shape assessment
in the other grade levels, they could implement other assessment programs—
which need not take the form of standardized exams—to ensure students would
be ready to take the exams in the tested grade levels.

The length of the project should be no more than five years. Notably,
planning for the ill-fated California Learning Assessment System took roughly
two years. A two-year planning period along with a three-year
implementation period should yield results about the success of the new plan.
Congress could extend the experiment period as needed.

2. Building Capacity Within the U.S. Department of Education to Facilitate
   Collaboration Among the States

Next, the experimentalist project requires a federal institution that can
facilitate information sharing among the states. The National Center for
Education Evaluation and Regional Assistance (NCEERA), housed within the
Department of Education’s Institute for Education Sciences, would be a likely
candidate for this function. NCEERA evaluates federally funded education
programs and is charged with disseminating its reports to “encourage the use of
scientifically valid education research and evaluation throughout the United
States.” Congress created NCEERA under the Education Sciences Reform
Act of 2002, with the intention for NCEERA to “provide national leadership
in expanding fundamental knowledge and understanding of education.”

150. See Michael W. Kirst & Christopher Mazzeo, The Rise, Fall and Rise of State
contentdelivery/servlet/ERICServlet?accno=ED397133.
§ 9511 (2006)).
NCEERA’s structure and mission would need to be adapted to accommodate the need for timely information for the assessment project. To facilitate this adaptation, NCEERA could partner with a national organization that has already completed research in this field, such as the National Governors Association (NGA).

Although Congress could charge NCEERA with disseminating information to the states, it would likely prove too slow to provide the fast-paced information states would need to conduct their policy experiments. A recent study showed that finished reports took an average of twenty-eight weeks to emerge from the NCEERA review process.\(^\text{154}\) One commentator suggested that the delay may be attributable to “political reasons,”\(^\text{155}\) which, if true, would undermine the credibility of the research process. Policymakers working to improve their assessment systems based upon experiences in other states would need access to information more quickly, and, to preserve the integrity of the process, the information cannot be tainted by allegations that it passed through political filters.

The NGA could assist NCEERA with disseminating information to the states more quickly.\(^\text{156}\) National organizations are seen as providing “more comprehensive and timelier” information than other sources,\(^\text{157}\) and, instead of reinventing the wheel, NCEERA could rely on work already done by the NGA. Indeed, national organizations, such as the NGA, already have credibility with policymakers who often turn to them for information.\(^\text{158}\) NGA would be a good candidate for such a partnership because of its “Center for Best Practices,” which collects information on the states’ policy innovations,\(^\text{159}\) and because of NGA’s experience in education reform.\(^\text{160}\) Beyond the information benefits, partnering with an organization like the NGA could yield political benefits by providing the support of a nonpartisan group for federal policies.

\(^{154}\) Debra Viadero, Institute of Education Sciences’ Board Calls Agency’s Past 5 Years a Success, EDUC. WEEK, Dec. 3, 2008, at 10.

\(^{155}\) Id.


\(^{157}\) KARCH, supra note 134, at 122.

\(^{158}\) See id. at 106-07.

\(^{159}\) Id. at 123; see also Nat’l Governors Ass’n Ctr. for Best Practices, Services for Governors and Their Staff, http://www.nga.org/cda/files/cbpbrochure.pdf (last visited Feb. 23, 2010).

3. Establish an Independent Commission to Evaluate States’ Best Practices and Implement a National Education Policy

Finally, the assessment project requires a body to promulgate regulations based upon the states’ best practices in order to implement the ultimate goal: a national assessment and accountability plan. The intent would be to ratchet up national regulation as best practices become apparent and to prevent states from backsliding during the assessment reform effort. Implementation of the national plan would end the experimentalist project.

I propose the creation of an independent regulatory commission that would implement the national assessment plan. The commission would be housed within the Department of Education. The commission would function similarly to other quasi-independent commissions, such as the Federal Aviation Administration (FAA), in that it would have power to promulgate regulations on its own but would function under the sphere of influence of a cabinet department.

The lightning-rod nature of education reform necessitates the political insulation that accompanies independent commissions. Take for example the California Learning Assessment System (CLAS) and its short, but tumultuous life. The CLAS assessed students through means other than multiple-choice exams. For example, to assess reading ability, the assessment required students both to participate in a small-group discussion and to write a response.\textsuperscript{161} To evaluate writing, the assessment required students to respond to texts chosen by teachers.\textsuperscript{162} The CLAS also relied less on the multiple-choice format in mathematics assessments.\textsuperscript{163} A cadre of teachers graded the exams during the summer.\textsuperscript{164} The hope for the CLAS was to create an assessment “developed by teachers and scored by teachers, under the leadership of teachers.”\textsuperscript{165} The new program, however, met a swift end. One year after its first administration in 1993, then-Governor Pete Wilson effectively ended the program by vetoing legislation that would provide future funding.\textsuperscript{166}

Politics played a substantial role in Wilson’s veto. Wilson, a Republican, faced reelection that year, and conservative and religious organizations deplored the CLAS.\textsuperscript{167} The groups complained that the tests asked improper

\textsuperscript{162} Id. at 17.
\textsuperscript{164} Dudley, supra note 161, at 17.
\textsuperscript{165} Id.
\textsuperscript{166} See Merl & Ingram, supra note 25 (reporting on then-Governor Wilson’s veto of funding for the California Learning Assessment System, which relied less on multiple-choice exams).
\textsuperscript{167} See Jean Merl, Furor Builds over Methods Used in Exams, L.A. TIMES (ORANGE COUNTY EDITION), May 6, 1994, at A1.
questions and introduced students to controversial topics. In the words of a conservative group, the test represented an "effort[] by the state and federal governments to usurp the fundamental rights" of parents. In the end, praise for the system's innovation in assessing students could not save the CLAS.

Although the CLAS also suffered from technical deficiencies, politics were central in its demise. Undoubtedly, national assessment reform will be unable to escape political influences. But lessons they learn from the CLAS can help future reformers avoid making past mistakes. Additionally, an independent commission can insulate assessment reform to prevent electoral politics—such as those seen around the CLAS and the 1994 California gubernatorial election—from bringing reform to an early end.

The membership composition of an education commission should be crafted to address political needs, which the developers of the CLAS failed to take into account. Such careful crafting is important, particularly if control over national assessment will be given over to an independent commission. The former National Education Goals Panel (NEGP) provides a model for membership composition that takes into account political considerations. The NEGP was comprised of eighteen members: two members appointed by the president, eight governors appointed by the NGA, four members of Congress, and four state legislators appointed by the National Conference of State Legislators. The assessment commission could be composed similarly, though some provision should be made to include members from a teachers' group (e.g., one member appointed by the National Education Association and another appointed by the American Federation of Teachers). To account for the problems with the CLAS, members from various backgrounds should be involved in the system.

Despite membership requirements that account for various constituencies, this proposal for an independent commission will likely be controversial. The commission's political insulation (and isolation) serves as both a benefit and a liability. Two options can make the proposal more politically palatable. First, the commission's charge can be narrowed to give it jurisdiction over creating a national assessment and accountability scheme. Once this purpose is met, the commission could be discharged. Second, regulations promulgated by the panel could be subject to existing statutory rulemaking protections. The Administrative Procedure Act (APA) would then govern the rulemaking

168. Id.
169. Id.
171. Id.
172. See Kirst & Mazzo, supra note 150, at 8-11.
173. Id. at 4 ("The designers of the CLAS items had not included potential critics—those that represented traditional religious and conservative groups. The specific wording of the questions had not been checked for possible objections with audience who might object.").
procedure, providing public participation in the process through notice-and-comment rulemaking and a structure for judicial review.174 Moreover, once final rules are promulgated, Congress could yet overrule the new regulations through its powers under the Congressional Review Act.175

The commission’s independence, however, is vital. Because this would be a multi-year endeavor, spanning congressional sessions and perhaps presidential terms, any student assessment reform could benefit from political insulation.

C. Race to the Top Program

Indeed, the Obama administration has adopted components of the proposal outlined above, but the program is intended to be temporary.

The Race to the Top program, funded by the American Recovery and Reinvestment Act of 2009, provides states with the opportunity to seek federal grant money for education.176 States may apply for grants to, among other things, improve equity in teacher distribution, develop longitudinal student data systems, and improve state academic content standards.177 A portion of the fund is devoted specifically to the development of new student assessments, which would be adopted by groups of states.178 Mindful of the ire that the specter of common assessments often raise, Secretary Duncan made clear that the states will “call the shots” with regard to assessment reform.179

Only consortia of states will be eligible to seek funding for assessment grants.180 High priority will be given to a consortium that encompasses a “majority of the States in the country.”181 The Department of Education has placed a priority on developing cooperation among the states.182 The Department notes that “the larger the number of States within a consortium, the greater the benefits and potential impact.”183 But projects funded by Race to the Top may last no more than forty-eight months.184 For such a program to have a pronounced effect, it should persist for more than two years.

Although it offers grant money to states to innovate in education and requires states to report annually to the Department of Education, Race to the Top provides no mechanism through which the states can share ideas and compare results. And even though the program will undoubtedly generate new

175. See id. §§ 801–08.
177. Id. § 14005(d)(2)–(4), 123 Stat. at 283.
179. Id.
181. Id.
182. Id. at 59,729.
183. Id. at 59,735.
184. Id. at 59,840.
ideas in education, there is no mechanism through which those ideas could be implemented nationally. Nevertheless, Race to the Top is a positive step towards creating a stronger relationship between Washington and the states. At the time this Comment went to print, the initial effects of Race to the Top were beginning to emerge and the Obama Administration had recently proposed policies intended to supplant those established under No Child Left Behind.\textsuperscript{185} A preliminary, limited review of these results and proposals suggests that significant changes in public education could occur within the next several years.

CONCLUSION

As a fourth-grade teacher in a California public school, I endeavored to balance the pressures of testing with my desire to give my students a well-rounded education. Although my school focused heavily on mathematics and English instruction, we worked to integrate social studies and science into my students' school day. We also managed to include arts instruction, physical education, and field trips.

But it was not easy. With only about six hours in the school day, our coverage of non-tested subjects could have been better, and the ever-present focus on making the NCLB-mandated AYP kept the faculty and students on their toes.

NCLB changed schools through its mandates for content standards and accountability. Given the resistance to federal education mandates during the 1990s, that all fifty states now have content standards and accountability systems as required by NCLB should be viewed with some incredulity. But despite its progress, NCLB remains imperfect. Its deference to local control, which looks increasingly anachronistic given students' need to compete in a global economy, creates inconsistencies among the states and hampers efforts to genuinely raise student achievement.

Assessment and accountability reform is ripe for federal action for several reasons. First, for inherent fairness concerns, a national accountability scheme cannot be based on yardsticks that vary state by state. Second, an increasing number of business and nongovernmental organizations are demanding that we assess students for workplace skills that will make our future workforce competitive in the international workplace. Finally, by relying solely on standardized tests, we ignore the inherent technical problems with making conclusions about student and school performance based on test scores.

The combined experiences of the states, both past and future, can guide the development of a national assessment and accountability system, and the federal government, through its resources and institutions, can coordinate the

interstate collaboration for this task. The procedural precepts of democratic experimentalism can provide the infrastructure for this new federal-state relationship.