The Problem of Capacity in the (Re)Design of Educational Accountability Systems

Richard F. Elmore
Graduate School of Education
Harvard University
October 2006

Accountability in education requires a complex mix of pressure and support. On the pressure side, accountability systems operate by surrounding schools with systems of standards, measures, targets, and sanctions designed to move performance in a particular direction. On the support side, accountability systems succeed to the degree that they stimulate demand for new knowledge about how to meet performance targets and generate the supply of knowledge that meets that demand. In this sense, educational accountability systems are special case of the more general class of regulatory policies. In general, regulatory policies try to get individuals or institutions to do something they might not otherwise do in the absence of external pressure or coercion. Such is the case with educational accountability systems, but with an important difference. The difference is that educational accountability systems are often—one might say usually—asking schools to do things they do not know how to do. That is, in order to meet the performance targets specified by accountability systems, schools and school systems must engage in classroom practices, develop organizational structures and routines, and manage resources in ways that are unfamiliar to them, and that require significant investments in learning. Sometimes the knowledge and skill necessary to meet performance targets is resident in the organization, or is easily acquired using the existing

---

1 This paper was prepared for the conference, “Examining America’s Commitment to Closing Achievement Gaps,” sponsoring by the Campaign for Educational Equity, Teachers College, Columbia University, November 13-14, 2006. Some of the research for this paper was supported by the Consortium for Policy Research in Education, under a grant from the U.S. Department of Education.
knowledge and skill of people in the organization. More often, educational accountability systems require a level of knowledge and skill that is not embedded in the organization, and the pressure of accountability activates a demand for new knowledge that may or may not be met by the environment in which schools operate. If accountability in education were simply a matter of compliance, then strong enforcement would produce performance. But accountability is also a matter of knowledge, which means that enforcement has to be counterbalanced with support. In Thomas Schelling's terms, this is the difference between "doing the right thing," and "knowing the right thing to do." ²

This paper is about the relationship between pressure and support—coercion and knowledge—in educational accountability systems. It provides a framework for understanding how accountability policies work, and a critique of existing policies, with special reference to No Child Left Behind. It focuses, in particular, on two meanings of the term "capacity" in educational accountability policies. The first is the capacity to enforce. Whatever the rhetoric that surrounds accountability policy, the underlying theory is that one level of government has the power to force another level of government (or a private firm) to do something that it would not otherwise do, or would not do in the same way, by administering sanctions, or withholding valued resources. The credibility of regulatory force depends on at least three factors: (1) the capacity of the regulatory agency to enforce; (2) the feasibility of the actions required by the regulations; and (3) the consent of the targets of regulations to be regulated. We will see that federal education policy in general, and No Child Left Behind in particular, presents an unusual set of problems around regulatory capacity.

The second meaning of capacity in accountability policy is the capacity to perform. Educational accountability policies require schools and schools systems not just to comply with regulatory requirements—of which there are many—but also to produce performance of a particular kind, according to a particular metric, over a specified period of time. No Child Left Behind requires schools and school systems to improve the performance of

---

students across grade levels in core academic subjects against a state-specified standard in such a way that all students meet a standard of proficiency by the year 2014. The capacity to produce performance is definitely not the same thing as the capacity to comply. Producing performance, as noted above, requires the development and deployment of knowledge and skill. What this capacity looks like and how it is formed are factors critical to the success of accountability policies.

The Capacity to Enforce

Federal Policy: Borrowing Capacity and Regulatory Drift

No Child Left Behind (NCLB) signals a major break with past federal policy. It constitutes a federal preemption of accountability policy, which had previously been for all practical purposes entirely a state function. Prior to NCLB, the federal government’s role in accountability was marginal at best. The 1997 amendments to the Elementary and Secondary Education Act, which were designed to situate the federal government in a more powerful position around state and local accountability systems, had failed to produce their intended effect. The governors and state legislatures had taken the initiative in forming accountability policy. All fifty states had developed and implemented accountability systems in collaboration with each other, and these systems, as is usually the case with the “laboratory of federalism,” embodied wide variations around the central elements of accountability policy. States used a variety of different types of tests, sanctions varied widely, stakes were distributed among key actors differently, and success and failure were defined in different ways. NCLB, in effect, preempted the states’ initiative. While the rhetoric around NCLB carefully asserts the primary role of states in accountability policy, the reality is considerably different. NCLB sets fixed parameters on state accountability systems that previously did not exist in federal policy. These requirements dramatically reduce the range of variation among state policy that previously existed: Annual testing at fixed grade levels, for example, limits the type of tests that can be used—more ambitious criterion-referenced tests are too expensive to administer and score for every student on an annual basis. The fixed schedule of sanctions under NCLB—schools progress through uniform stages of sanctions, on a uniform schedule, as they fail to meet performance standards—limits states’ and localities’
flexibility in identifying failing schools and crafting remedies for them. The schedule of performance gains prescribed by the law—every student at proficiency by 2014—requires all states to define the required rate of improvement in the same way, and introduces strong incentives for states to lower their definitions of proficiency. The requirement that parents receive notification of low performance in their schools and access to alternative services prescribes the same remedy for low performance in all instances, limiting flexibility in responding to differences among schools. The requirement that every child should have a "highly qualified" teacher attempts to create uniformity in what is by definition a widely varying patchwork of regional labor markets, each with its own particular characteristics of supply and demand. Each of these elements, and many more, in NCLB is part of a dance of federalism, in which the federal government effectively preempts state and local accountability policy, while at the same time arguing that what it is actually doing is participating in a partnership in which states and localities exercise control over the specifics of their accountability systems. However one characterizes this dance, it represents a dramatic shift in the relationship between the federal government, states, and localities in the governance and control of education.

Underlying this dance of federalism is a long-term, very stable reality: The federal government has little independent capacity to produce educational outcomes at the state and local level. Its power is effectively limited to its ability to capture and redirect state and local capacity. The federal government’s operating budget—as opposed to its grant-in-aid budget—is small relative to state and local operating budgets. Hence, its capacity to engage in direct oversight and enforcement of regulatory requirements is severely limited. Federal expenditures in the K-12 sector are a tiny fraction of total expenditures in the sector. Hence, its capacity to exercise influence over state and local decisions directly, through the use of financial controls, is limited. The way federal influence works is highly specialized. While federal expenditures are a small proportion of education expenditures in general, they are a significantly larger proportion of expenditures in urban school systems, and in suburban systems with increasingly diversifying populations. Federal funds also constitute the major source of funding for most state educational agencies, which are, in effect if not in principle, wholly-owned subsidiaries of the
federal government. Federal, state and local relations in the K-12 sector, then, are characterized by a long-term dynamic that Paul Manna calls “borrowing strength”—what I will call “borrowing capacity.”

Manna’s framework predicts that policy relations between the federal government and the states hinge on three factors: (1) the presence of policy entrepreneurs who capitalize on latent issues and turn them into policy problems; (2) the license to act, or the presence of a statutory or constitutional rationale for intervention; and (3) the capacity to act or the “human, budgetary, and institutional resources to” to implement its claim to license to act. When policy entrepreneurs operate in a policy domain where the federal government has high license and high capacity, they are likely to intervene without regard the capacity of states or localities. Lacking high license and capacity, federal policy entrepreneurs are likely to borrow strength from states and localities. Having made the decision to borrow strength, the success of federal policy depends on the accuracy of policy entrepreneurs’ assessment of the license and capacity of state and local agencies in the domain where they have chosen to act. As the limits of state and local license and capacity become apparent, Manna’s theory predicts, the federal government will negotiate with states and localities to modify the demands of federal policy.

What does “borrowing strength” or “borrowing capacity” actually look like? First, federal revenue constitutes a very large share of the operating revenue of state departments of education—anywhere from 10% to 50%, depending on the composition of the student population and the mix of federal programs that states are involved in. The combination of federal revenue and the regulatory load of federal policy means that state departments of education in states with large populations of children in poverty are close to being wholly-owned subsidiaries of the federal government, at least for accountability purposes. It also means that state department organization and state capacity tends to mirror federal organization and capacity, heavily driven in the direction of regulatory oversight and away from investments in knowledge and skill at the local level.

4 Ibid., 28-33, 31.
5 Ibid., 87, Table 5.1.
In California, for example, federal revenue from all sources accounts for about $7.5 billion of a $66 billion budget. About $41 billion dollars of funding for elementary and secondary schools comes from the state general fund; about $13 billion from local property taxes. So, in effect, federal regulatory leverage in California operates on about 9% of the state’s total educational expenditures. But it capitalizes on the fact that the state provides about 76% of local revenues for education. In Los Angeles, Title 1 accounts for about $820 million of a $13 billion education budget (all federal programs account for about $1.2 billion). So federal regulatory leverage in Los Angeles operates on a ratio of about 6%. These ratios will vary across states, but the underlying principle is clear: Federal policy operates on a relatively small margin, using its regulatory power to influence the use of large amounts of state and local capacity, relying on influencing the flow of revenue from state and local sources as a means of control.

Within the K-12 domain, the federal government lacks both the financial and operational capacity to enact its policy goals independently; federal policy works, insofar as it works at all, by the federal government using its limited leverage in a concentrated way to “borrow” the capacity of states and localities for its own ends. The limited fiscal and operational capacity of the federal government means that borrowing capacity has to be, of necessity, a primarily regulatory activity. That is, the federal government can transfer money to states and localities, but the amount of money it has at its disposal is not, in itself, large enough to influence state and local decisions, much less to fill the gap in organizational and individual capacity within schools required to meet its goals. Its influence depends on its capacity to use money as the basis for a regulatory control. Federal influence depends heavily on the capacity of the federal government to use the threat of regulatory enforcement to cause states and localities to do what it wants them to do. Whether it wants to or not, whether its rhetoric says so or not, the federal government’s intergovernmental influence is solely a function of its regulatory authority.

---

6 Manna, *School’s In*, 106ff, 142ff.
Whatever the intent of federal policy, then, its effect will always be primarily regulatory. I call this tendency “regulatory drift.” The broad goals of No Child Left Behind may be described in terms of improvements in the quality of learning for children, or, by a more or less calculated appropriation of civil rights rhetoric, as providing equal access to quality education for poor children and children of color. In effect, though, given the realities of federal influence, these broad goals are only as good as the force of regulatory authority that the federal government exercises over state and local decisions.

In the current U.S. Department of Education budget of about $7 billion for NCLB and related programs, about $5 billion is in formula grants to states and localities through various provisions in Title 1. Of the remaining $2 billion or so, the largest share—a little over $1 billion—is in Reading First, a highly regulated categorical grant program, in which the federal government determines which reading curricula will be used by grantees. The vast majority of funds that flow from the federal level to states and localities are subject to the regulatory regime of NCLB. The federal government makes minimal investment outside the regulatory structure in raising the level of capacity of state and local agencies to meet the regulatory requirements, and, when it does—as in the case of Reading First—it essentially runs a capacity-building program as a highly regulatory activity.

In mid-2006, USDOE Secretary Margaret Spellings released a set of proposals designed to respond to critics of NCLB, noting how important it was for policymakers to “listen carefully to the feedback from those on the ground so that they can better understand how the law translates from paper into action,” and acknowledging “the need to give states some alternatives in implementation” of NCLB. The ensuing proposals deal exclusively with modifications of the testing provisions of NCL, allowing states to modify the AYP requirements for students with “persistent academic disabilities.” The Secretary directed that $14 million be reprogrammed to deal exclusively with the assessment issues stemming from this change in policy, another $3 million in grants to states to improve state assessments for students with “persistent academic disabilities,” and another $6 million dollars to “improve accountability” and “track

---

changes” in the effects of assessment shifts on students. Improvements in NCLB, from the federal perspective, are clearly equated with changes in testing, not with resources directed at improvement of instructional practice.

Regulatory drift is not solely a characteristic of federal policy. It occurs, as Manna’s theory suggests, wherever the goals of governmental institutions exceed their capacity to meet those goals. In accountability policy, the dynamic of regulatory drift is played out across all levels of government. Federal policy drifts toward regulatory control of state and local agencies. State administration of federal policy drifts toward regulatory control of local agencies. Local administration of federal and state policy drifts toward regulatory control of schools. In each instance, agencies think they are acting in the interests of their clients—the children. In each instance, they think they are adding value to the service—quality teaching and learning. In each instance, the drift toward reliance on regulatory control stems from the fact that agencies are doing what they know how to do. Bureaucracies know how to make and enforce rules. They have great difficulty doing anything else. While it is not impossible for them to engage in other kinds of activities, non-regulatory activities require them to move out of the regulatory zone, and in doing so, to do things they fundamentally don’t know how to do, or don’t know how to do well.

One thing educational bureaucracies know how to do well is administer tests. Testing is relatively cheap as an instrument of regulatory control. Procedures for test administration and for compiling and disseminating test results can be stated in relatively clear regulatory language and implemented with relatively clear procedural controls. The process of testing can be run using familiar regulatory routines—oversight, inspection, enforcement, sanctions. Since bureaucracies tend to do more of what they know how to do, and they tend to avoid doing those things they do not do well, it is not surprising that accountability has increasingly come to be equated with testing. State and local educational agencies are thought to be “accountable” when they are administering tests, publicizing test results, analyzing test scores, developing and administering their own tests, helping school personnel analyze and use test results, and so on. In education,

---

then, regulatory drift is synonymous with a drift toward increasing use of tests as instruments of control.

**Regulatory Enforcement as Strategic Interaction**

Every regulatory system involves a critical trade-off between standards and enforcement: The more stringent the standard relative to existing performance, the greater the likelihood of failure to meet the standard, and the greater the cost to the regulatory agency of inducing compliance or performance. The lower the standard relative to existing performance, the greater the likelihood of compliance, and the lower the costs of enforcement. In real life, regulatory agencies play this game in a more or less deliberately deceptive way: They tend to set standards well above the level they know they can enforce and bet that by skillful use of enforcement—e.g., by calling attention to “worst cases”—they can induce a higher level of compliance than they actually have the resources to manage. This problem is known in the regulatory theory literature as “optimal standards with incomplete enforcement.”

The regulatory regime of No Child Left Behind creates a framework for identifying low-performing schools that applies to all states and localities for. Schools that fail to meet the increment of performance required by Annual Yearly Progress (AYP) advance through three distinct stages of regulatory sanctions: school improvement, corrective action, and restructuring. Each stage carries with it a particular set of required actions, at the school and school system levels, and a specific period of time, typically two years, within which the actions have to be completed and the next level of performance under AYP has to be met. States can choose how they define progress within the AYP framework, providing that all students reach

---

10 W. Kip Viscusi and Richard Zeckhauser, “Optimal Standards with Incomplete Enforcement,” Public Policy, Vol. 27, no. 4, 437-456; Carol Adaire Jones, “Standard Setting with Incomplete Enforcement Revisited,” Journal of Policy Analysis and Management, Vol. 8, no. 1 (1989), 72-87. Viscusi and Zeckhauser demonstrate that, with lump-sum penalties (the same penalty for all violators), regulators tend to set standards too high. Jones demonstrates that, with graduated penalties (penalties that increase with the seriousness of the violation), the Viscusi/Zeckhauser analysis probably doesn’t hold, and regulators should set the standard at the highest possible level that represents the true value to society of the violation.
the “proficient” level by 2014. The performance standard required for AYP must be met by all students in a federally-prescribed set of categories—limited English-speaking, special education, low-income, etc. At each stage, the sanctions that apply to schools for failing to meet AYP are exclusively procedural: school planning, notification of parents, provision of funding for access by students to externally-provided supplemental services, restructuring the school’s organization, closing the school and reopening it under different auspices. What schools need to know in order to improve, and who is responsible for assuring that knowledge is present, is not specified in the law.

From a regulatory perspective, there are a number of notable features of NCLB. First, there is no empirical basis for the AYP requirement. There is no body of evidence, for example, that tells us what a reasonable period of time might be for a school to progress from one level of performance to another. Nor do we do we know, within schools, what the expected increment of progress might be for different groups of students over a given period of time. There was no evidentiary basis for the sanctions prescribed by NCLB, or the schedule of progressive sanctions. Standard-setting and enforcement schedules in environmental policy, for example, involve vigorous debate around the evidence, with contending parties offering differing interpretations and regulators accepting some obligation to base their final decisions on a body of evidence about the benefits and costs of the regulations to the various parties and feasible compliance schedules. No such debate has occurred with NCLB. No evidentiary basis exists for such a debate.

Because there is no evidentiary basis for the AYP requirement, there is no mechanism in the law, or in the regulatory structure following from the law, for adjusting the number of schools identified as failing to the resources available to remediate their failure. In other words, there is no basis for making a strategic judgment, at any level of the system, relating the level of resources available for enforcement or remediation of performance to the actual caseload of schools that are in need of improvement under AYP. States can adjust their proficiency standards downward, at the risk of incurring federal sanctions, but otherwise there is no mechanism in the system to equilibrate the regulatory load produced by AYP to the resources available for school improvement.
If success in AYP were simply a matter of compliance—doing something people already know how to do, but might choose not to do when they consider the costs of doing it—this condition would not be a serious matter. Federal and state regulators would simply do what rational regulators normally do under these circumstances: They would engage in strategically calculated, high-visibility, selective enforcement, designed to signal to schools and districts that they run the risk of sanctions for failure to comply. This strategy keeps enforcement costs under control and compliance at an optimal level. Schools and districts would then calculate the probability of getting caught and come to their own conclusions about whether and how to comply—just like polluting firms in environmental regulation.

But AYP poses a different kind of problem of regulatory compliance. Most schools get classified as failing under AYP because they don’t know what to do to get better. No amount of regulatory enforcement, no matter how cleverly contrived, will rectify this condition. Surrounding failing schools with procedural requirements, as NCLB does, doesn’t solve the knowledge problem, and indeed, can make it worse, by focusing failing schools’ on compliance rituals—planning, consulting, issuing notices, filing reports—rather than addressing the lack of knowledge and skill that produced the problem in the first place. Furthermore, failing schools will not get better simply by pulling up their socks and doing what they are supposed to do. They can pull up their socks, but they will probably do is a more systematic version of what they are already doing. Schools get better as a consequence of bringing new knowledge and skill into their practice—individually and collectively—not by doing what other people tell them to do. Furthermore, as accountability systems mature and as regulatory drift progresses, regulators focus more and more on tested performance, and less and less on the instructional conditions that produce that performance.

Studies of the effect of regulatory sanctions on low-performing schools have produced uniformly negative and cautionary findings. Teachers and principals generally acknowledge and internalize the negative signals the accountability system is sending them about their performance. That is, they know they are low-performing. The regulatory sanctions themselves generally have either very weak or non-existent effects on the ability of schools, or the individuals who work in them, to improve their performance, even in circumstances where there are
well-developed intervention strategies to improve failing schools. Under the pressure of sanctions, teachers tended to do what they knew how to do—largely to teach basic skills-oriented content with low-level pedagogy—and what they knew how to do was generally mismatched with the requirements of the accountability systems in which they were operating. The accountability systems "demanded that "teachers [become] motivated to learn and become proactive. In the observed classrooms, however, such learning was widely absent, and instructional change stalled." 11

This combination of circumstances—a regulatory regime that is disconnected from actual evidence of the problem it is trying to solve, the lack of a mechanism for equilibrating the number of failing schools with the resources available to deal with them, and the lack of fit between procedural remedies and the causes of failure in schools—creates the situation we are currently in. The number of schools that are in various stages of sanction under NCLB bears no relationship to the capacity of the federal, state, or local officials to remedy their performance. And the requirements for the remedy, as well as the resources entailed in the remedy, are largely opaque and inscrutable to those who are responsible for it.

In the best of all worlds, this problem would be self-correcting. The federal government would revisit AYP and try to establish a reasonable basis for its regulatory requirements. The states would revisit their responses to the AYP requirement based on their capacity, and the capacity of local districts to assist schools in meeting performance requirements, and so forth. But the particular incentive structure of NCLB makes this kind of self-correction unlikely. The federal government has no incentive to address the underlying conditions that produce the number of failing schools under AYP, because the federal government does not have to bear directly any of the costs of failure. Under the borrowing strength theory, the federal government simply transfers money from one level of government to another, creates the regulatory regime that produces failing schools, and passes the costs

of rectifying failure along with the responsibility for regulatory enforcement on to states and localities. Likewise states take a primarily regulatory posture toward school failure, and pass along the costs to districts and schools. Some districts recognize that they are bearing the costs of performance, some don’t. In either situation, district and school people end up with the knowledge and skill problem in their laps, surrounded by a regulatory policy that does nothing to address the fundamental gap in knowledge and skill that created the performance problem in the first place. So a critical part of the feedback loop that would lead to self-correction is missing.

Furthermore, the political calculus at the federal level does not lead to self-correction. If NCLB succeeds in raising student performance and reducing the number of failing schools, then federal policy makers will generate electoral credit by claiming responsibility for success. If NCLB results in no change in student performance and the number of failing schools, or an increase in failing students and schools, federal policy makers claim electoral credit by blaming the system that produces this result.\(^\text{12}\)

In Massachusetts, the total number of schools identified for improvement under various stages of NCLB in 2006 is 617 (316 for aggregate performance, 301 for subgroup performance), up from 420 in 2005. Among the newly-identified schools, 57 are in the most severe category—restructuring, based on aggregate performance—up from 30 the year before. Between 2005 and 2006, 45 schools were removed from the list of schools needing improvement. Ninety of Boston’s 145 schools are identified for improvement under NCLB—thirteen in restructuring. Of the 115 lowest performing schools, 80% are in eight urban districts. This pattern—dramatic increases in schools entering various stages of improvement under NCLB, slow progress in schools leaving, and disproportionate numbers of school in need of improvement in the most populous urban areas—is consistent across the country.\(^\text{13}\)

---

\(^{12}\) See Richard Elmore, “The Politics of Accountability” (in process). For an account of the specific political process that produces this result, see: Elizabeth DeBray, Politics, Ideology, and Education: Federal Policy During the Clinton and Bush Administrations (New York: Teachers College Press, 2006).

\(^{13}\) “617 Schools Identified as Needing Improvement,” Boston.com; Jimmy Kim and Gail Sunderman, Large Mandates and Limited Resources: State Responses to the No Child Left
Massachusetts allocates about $5 million for a program designed to provide regulatory oversight, intervention, and support to failing schools. MassInsight, a non-partisan, not-for-profit, non-governmental organization that also provides support to failing schools, estimated in 2005, when there were 400-plus schools in need of improvement, that the state should be spending at least $15 million—proportionately increased to meet the 2006 demand, that number would over $20 million. Clearly, there is no relationship between the regulatory caseload spun off by the AYP requirement and the capacity of states and localities to meet the demand for support for failing schools. Equally clearly, there is no mechanism in either the policy or politics of NCLB for equilibrating capacity and demand.

The characteristic that distinguishes NCLB from a straightforward regulatory policy is that it is a regulatory policy nested within a grant-in-aid policy. The federal government transfers money to states and localities on the expectation that it will be used to benefit children, and, more importantly, that it will be used to drive state and local accountability policy in a given direction according to certain parameters set down in the policy. The costs of compliance at the state and local levels can be transferred to the grant-in-aid program, they can be borne out of general revenue by states and localities, or they can be met by some combination of the two. Federal and state bureaucrats tend to see the use of money for regulatory compliance purposes, from whatever source, as part of the process of holding schools accountable for performance. People in schools, and some people in local districts, tend to view the use of money for compliance purposes as a dead-weight loss—that is, money used for compliance is money that is not available for use in schools and classrooms to improve instruction. One of the problems with NCLB as a regulatory policy is that is it never clear how much of the grant-in-aid money is used for compliance, how much it actually costs state and local agencies out of general revenue to meet the regulatory requirements of NCLB, and how much federal revenue actually goes to support improvement of instruction in classrooms for children.

Such is the consequence of borrowed capacity. It is in the interest of the federal government to bury

_Behind Act and Implications for Accountability_ (Cambridge, MA: Harvard Civil Rights Project, February 2004).
enforcement and compliance costs in the transfer of revenue from one level of government to another, and to avoid acknowledging the compliance costs to state and local jurisdictions out of their own revenue sources, in order to make it difficult for states and localities to judge whether the benefits of federal grant-in-aid money actually exceed the costs.

**Moral Suasion, or “The Devil Made Me Do It”**

One of the chief functions of regulation is to force us to do things we might not otherwise be inclined to do, whether we want to do them or not. This is the “doing the right thing” half of Schelling’s formula. This view of policy lies behind the appropriation of civil rights rhetoric to justify the regulatory role of the federal government in NCLB. Putting the federal government in a regulatory role not only nationalizes educational policy in an arena that had previously been the primary preserve of states and localities, it creates an economy of blame that allows state and local administrators to minimize their own role, and their accountability to state and local institutions, by transferring blame to the federal government. This is the reciprocal side of the blame-shifting that occurs when federal policy makers blame the lack of success of the law on the failure of local institutions, rather than on defects in the law itself, and their own accountability for those defects. The superintendent of an urban school system said to me, “There are parts of this law that are killing us on the ground and making our work much more difficult. But I don’t want the law changed on my watch because it provides leverage that I wouldn’t otherwise have over schools. Wait another couple of years and then fix it.” Many superintendents of high impact school systems are, not surprisingly, unwilling to express opposition to the law, even when they feel it is dysfunctional because they do not want to be seen as being on the wrong side of a civil rights issue, and because the law provides them with a ready-made pretext for decisions at the local level that they might otherwise not have the stamina or the political support to make. “The devil made me do it” is a useful tactic for bolstering leadership when the demands of the work exceed the skill and commitment of people on the ground.

The devil-made-me-do-it tactic works best when the person using it is predisposed to do what the policy requires and is looking for an external pretext to do it. Advocates of NCLB point to the support of leading superintendents and principals as evidence that the law
“works,” strategically avoiding the population of school leaders who are unconvinced of the merits of the law, or more likely, simply unprepared to translate the law into any coherent improvement strategy in their schools or districts. By definition, these leaders constitute the vast majority of people responding to NCLB.

When I work in the rarified atmosphere of school leaders who are committed to the values behind NCLB, our discussions operate at a relatively sophisticated strategic and tactical level. These practitioners understand how to use external policy as a lever to improve schools. The problems of school improvement at scale are difficult enough under these circumstances, but at least one can assume some level of knowledge about how accountability can work to reinforce school improvement. When I occasionally wander outside this rarified atmosphere of committed school leaders into a more random sample of practitioners, the absence of understanding of the relationship between the regulatory regime of NCLB and the work of school improvement is shocking and depressing. These administrators understand NCLB, insofar as they understand it at all, solely as a regulatory game, and their role as purely bureaucratic. If the feds beat on the state, and the state beats on them, their role is to beat on people in schools. “The test” is something that exists to be beaten, and any tactic that allows it to be beaten is a good one.

People who work in schools generally have little or no control over the learning of students, since learning is determined by factors in the lives of children outside of school. Therefore, the best tactic is to do the minimum necessary to comply with the regulatory requirements of the law until the system acknowledges that the accountability structure doesn’t work. In these situations, NCLB, not surprisingly, has little discernible effect on the attitudes and practices of administrators, and generally reinforces negative stereotypes of the effect of external policies on schools and classrooms.

The literature on bureaucratic responses to policy predicts that “bureaucrats”—defined as anyone in a position to exercise discretion in the implementation of a policy—can choose (1) to respond in the way they are supposed to (“work”), (2) to deflect the requirements of the policy and continue to do what they were doing (“shirk”), or (3) actively undermine the policy in accord with their own
interests ("sabotage"). The evidence of empirical research suggests that the factors that most influence the tendency of bureaucrats to "work" are the consistency of the policy's requirements with the preferences and skills of the individual ("functional" incentives) and the preferences and skills of the individual's peer group ("solidary" incentives). The factor that is consistently least likely to affect a preference for "work" is hierarchical supervision. In other words, vertical sanctions are relatively weak in the structuring the discretion of bureaucrats as compared with their own preferences and skills and those of their peers. Leverage over the performance of organizations is most likely to be effective when it focuses on shaping the knowledge, skill, and values of agents, rather than on strengthening the command and control functions of the system.

Much of the rhetoric of NCLB focuses on "doing the right thing" by children at risk of failure in school. The primary policy instrument is regulation of performance. The underlying theory of action is that educators will change their values in response to changes in external rewards and sanctions, and vertical controls, and that this change in values will result in changes in practice that, in turn, result in changes in performance. Regulation is predicated on the belief that people can be encouraged, persuaded, or shamed into doing the right thing by changing the external rewards and sanctions under which they operate. Performance-based regulation is designed to make the consequences of peoples' actions visible, therefore subject to public scrutiny, and useable as a source of encouragement, persuasion, or shame.

As with most regulatory regimes there is a kind of proceduralism that underlies this theory of action. If people behave according to procedures prescribed by the policy and incorporate the goals behind the policy into their belief systems, then they will learn to do the "right thing," and having learned to do the right thing, they will have the intended effect on student performance.

---

The main problem with this theory, as with most regulatory theories, is that the procedures embedded in the policy carry large capacity requirements for their successful implementation and do not operate in the same way on every organization. In schools and school systems that, for whatever reason, have acknowledged disparities in access and performance among students, and among schools with a system, making them go through the process of measuring and reporting performance creates a stimulus to examine their organization and practice. In schools and school systems with little or no awareness of disparities in student access and performance, and with little understanding of how their organization and practice have produced these disparities, the process of measuring performance has little effect. One of the first things that people report who work with schools on the use of data is how little the availability of data on student performance is related to its use, and how little the examination of student data leads directly to changes in practice. Performance-based accountability systems produce mountains of student performance data, most of which is never examined in any systematic way by people working in schools. The regulatory procedure can create the evidence, but it cannot create the predisposition or the capacity to use it, much less the knowledge of assessment necessary to account for its limitations.

So while regulatory policy can operate in powerful ways as a pretext for strategic action in systems that are inclined to act in accord with the goals of the policy, it is relatively powerless to create action where the predisposition doesn’t exist. Furthermore, as we shall see later the idea that people internalize the goals of a policy and then behave consistently with those goals is, at the very least, somewhat suspect. Moral suasion, whatever its attractiveness in principle, has limited traction on the ground.

The Capacity to Perform

Accountability as Institutional Response: Internal and External Accountability

---

If external pressure alone could produce performance in schools, then testing and sanctions would produce a more or less uniform, positive effect on student performance across all schools. In reality, the response is anything but uniform or positive. Uniform systems of testing and sanctions produce widely varying results across schools and school systems. Accountability, it seems, is not a matter of compliance; it is largely a matter of institutional response.

Institutional response, as a theoretical position, holds that organizations of all types, and schools in particular, vary along a number of dimensions that affect their responses to external forces in their environment: Some of these factors are objective—experience and expertise of teachers, material resources in classrooms, demographic characteristics of students, prior performance, mobility of student and teacher populations, etc. Some of these factors are less tangible, but at least as important—beliefs and expectations about students’ capacities to learn, knowledge of pedagogy and content, norms of individualism and collective work, skills in organizing and managing resources and people, etc. Institutional response theory posits that the effect of an external force on an organization, like a school, is a product of the rewards and sanctions embodied in that force and the internal features of the organization. Hence, application of a uniform set of regulatory rules and sanctions to a highly variable population of organizations produces a highly variable organizational response. This is how policies that are designed to produce greater uniformity in performance, like NCLB, often produce greater inequality, at least in the short term.

The regulatory mindset sees the world as a collection of organizations in various stages of compliance with the requirements of the law, and sees the purpose of policy as producing a more or less uniform response to a set of prescriptions. The institutional response perspective sees the world as a collection of organizations with widely varying characteristics affecting their ability to understand, interpret, and act on external pressures, and views policy as one of many external forces operating on organizations. The regulatory mindset sees the

---

organization as a receptacle for the demands of policy makers; the institutional response perspective sees organizations as active agents in determining their responses to various external demands, including specific policies. The regulatory mindset sees a single policy—a single set of rules and sanctions—as the central driver of organizational behavior; the institutional response perspective sees the effect of any given policy as determined by a signal-to-noise ratio of that policy among the other competing demands operating on the organization.

Advocates of NCLB regularly employ the conceit that before NCLB schools were “not accountable,” and with the advent of NCLB schools are now finally “accountable.” Policy makers and policy geeks generally like to imagine that the world begins (and possibly will end) with their ideas. This conceit contains two serious misconceptions. The first is a simple matter of history: Performance-based accountability in public education was a state and local invention, not a federal invention; NCLB preempted and nationalized a particular model of performance-based accountability, it did not invent the idea. The second misconception is more germane to the issue at hand. Institutional response theory does not assume that accountability begins with policy. In fact, all organizations, regardless of the policy environment in which they operate, have to solve the accountability problem—they have to decide to whom they are accountable for what, and how. Schools, consciously or not, have a solution to the accountability problem embedded in their organizational structures, routines, cultures that antedates the introduction of performance-based accountability in their external environment. So policy does not “introduce” or “create” accountability where none existed before; policy complicates, interacts with, and attempts to alter an existing accountability system.

A brief example will illustrate: The Cambridge, Massachusetts Public Schools, with whom I am currently working as a consultant, has, for the past 30 years or so had a “controlled choice” enrollment plan. Under the plan, parents were allowed to express a preference for any elementary school in the city, and places in schools were allocated based on a complex system. This plan introduced relatively tight relationships between parents—at least those who were active choosers—and school personnel; the

---

18 See Manna, School’s In, 86-116; DeBray, Politics, Ideology, and Education, 27ff.
plan also introduced significant competitive incentives among schools. Each school thought of itself as embodying a particular point of view about instruction and a particular set of relationships with parents and students. In other words, the plan created a very specific accountability structure. The advent of performance-based accountability in Massachusetts, in 1993, created serious conflicts with this embedded accountability system. It introduced sobering data about student performance in schools. It forced a common metric for measurement of student performance. And it eventually forced the district to develop a system-wide strategy for instructional improvement. Whatever one thinks about the merits of either accountability structure, it is important to acknowledge that the problem for Cambridge educators is not how to implement performance-based accountability. It is how to displace a deeply-seated existing accountability system that served certain interests extremely well with another accountability system, mandated by an external authority with very little legitimacy in the community. The problem was not how to make schools accountable. The problem was how to make them accountable for different things in a different way. Every school in Cambridge has had to go through this process, some more expeditiously and gracefully than others.

A principle of institutional response theory, then, is that uniform policies produce differential responses and producing more uniform results requires differential treatment. What determines the effect of an accountability system at the school level is not the policy itself but the interaction of the policy with the specific conditions inside the school. The same policy produces different responses in different settings. If the intent of policy is to produce greater uniformity in effects—this is called “equity” in the jargon of NCLB—then different organizations have to be treated differently. It is at this point that regulation loses its traction as a policy instrument. Regulatory policy is good at setting uniform standards for different organizations, not at adjusting the demands of policy to different settings. In fact, most of the dysfunctions of regulation arise from its incapacity to adjust to differences among organizations.

Our research on the specific conditions that determine a school’s response to external accountability pressure has
focused on a construct we call *internal accountability*. Briefly stated, internal accountability is the convergence between what individuals think they are responsible for, the expectations within the organization that express common norms and values, and the *processes and structures* by which people inside the organization account for their work. An organization with low internal accountability is one in which individual preferences account for most of what happens in the organization, collective norms and values exercise limited influence over individuals' work, and the processes by which people account for what they do are weak and infrequent. We call these organizations relatively *atomized*. An organization with high internal accountability is one in which there is a high convergence between what individuals say they are responsible and what the organization as a whole espouses as its values, and there are regular, visible processes by which people account for their work. We call these organizations relatively *coherent*.

Pushing hard with an external force—like testing and sanctions—on an atomized organization, our research suggests, does not make it a more coherent organization, at least in the short term. In fact, it often makes the organization more atomized and dysfunctional because people continue to do what they know how to do, which is exactly what produced the performance that got them in trouble in the first place. Pushing hard on a coherent organization, on the other hand, often makes it more coherent, at least in ways that respond to the external force. Another way of saying this is that schools generally respond to external pressure by doing what they already know how to do, or some modest extension of what they know how to do.

Schools with low internal accountability mostly know how to manage around the edges of instructional practice, preserving an essentially atomized organization. So they generally adopt responses that least challenge established instructional practice and that stress compliance with external directives. Teaching test items is one standard response. Teaching test items is something everyone can do, without incorporating any new knowledge into their practice, and it demonstrates compliance in a visible way.

---

19 Abelmann, Elmore, et. al, “When Accountability Knocks. . .”

But teaching test items doesn’t make the school a more coherent or effective organization around instructional practice. It is simply an accommodation to external pressure that leaves as much as possible of the existing organization intact.

Schools with high internal accountability generally have norms about what good teaching and learning look like. These norms may or may not be consistent with the messages they are getting from external tests and sanctions. Surprisingly, we did not find that it mattered a great deal whether instruction in high internal accountability schools was consistent or inconsistent with the norms embodied in the test. Because they knew how to manage instruction, they generally did well on the tests with a minimum of disruption to their internal processes. Their most visible accommodation to the testing regime was to spend a significant amount of time preparing students for the form of the questions, assuming that the students would know enough from their experience in school to handle the content. At the most, these schools would sometimes move the sequence of the content around to correspond more closely to the content of the test, but they did very little else to change their internal practices in response to the test.

In addition, for many nominally high performing schools with students of relatively high socio-economic status, the external test is not a serious accountability mechanism at all. These schools operate in a parallel accountability environment that stresses attainment, not performance on the state accountability test. That is, parents and school personnel focus on practices that situate their students favorably in a competitive attainment structure, leading to access to competitive institutions of higher education. This process begins early—typically in fourth or fifth grade—when school...

---

21 I have learned to use the term “nominally high performing schools” because my observations of classrooms in many—probably most—of these schools reveal that the instructional practice is typically mediocre or worse. Social capital—defined in this case as the advantages of social class and access to supplementary instruction—support a large part of what students learn in high socio-economic communities. Instruction in school apparently provides less. See Richard Elmore, “What (So-Called) Low Performing Schools Can Teach (So-Called) High Performing Schools,” Harvard Education Letter, September/October 2005.
personnel begin to make decisions about selective admission to high-track academic courses: “advanced” math, “honors” English, etc. There is usually no pedagogical rationale for these decisions, since privileged children can usually handle demanding academic work with skillful teaching. The main basis for these decisions is to signal to parents, and to their main accountability structure—institutions of higher education—that their schools are serious about academic work and that not everyone can succeed at “serious” academic work. For these schools, performance-based accountability is a purely prophylactic activity. They need to demonstrate a certain level of performance on the test in order to avoid public displeasure, and to justify the real estate prices that families have to bear in order to participate in the school system. The “real” accountability system in these schools, however, is not the performance-based one; it is the competitive attainment structure that they themselves design in response to the increasingly complex regime of competitive college admissions. It is not surprising that the accountability structure of attainment-oriented schools stresses access to high level content—numbers of advanced courses offered, for example—and success in college placement, not performance under the external accountability system. The challenge for attainment-oriented schools is how to create an internal accountability system that mirrors the competitive market for college admissions, not how to produce performance the regulatory system.

Internal accountability, then, plays a key role in determining schools’ responses to external testing and sanctions. Given this reality, it is easy to see how (a) the initial response within school systems to performance-based accountability systems might be to increase inequalities in performance among schools rather than to decrease them, and (b) how increasing the performance of school systems necessarily entails treating schools differently rather than engaging in equal enforcement of testing and sanctions. External accountability systems produce a range of responses, depending on the internal accountability of individual schools. Improving performance across a range of schools, each with a somewhat different internal accountability structure, requires differential treatment, rather than uniform enforcement. It is at this point that the incapacity of regulatory systems to produce uniform improvements in performance usually becomes obvious. Telling people to do the right thing doesn’t help them learn the right thing to do.
Institutional Capacity as Social Capital

I have often thought how different our accountability systems would be if policy makers, and policy analysts, were required to spend an entire day working in a school for every day they spend making policy for schools, or pontificating about how to fix what is wrong with schools. Observing classrooms, and the organizational and managerial work that surrounds them, is a sobering and disorienting experience. Hard-edged preconceptions about what is wrong with schools and how to fix them evaporate into a haze of difficult questions. It is much easier to prescribe from a distance. Lessons from research and from “best practice” that seem to fit perfectly in some settings collapse in a heap of steaming rubble in others. Improvement strategies that seem robust initially run out of steam and stall after a period of time. What seem like great ideas at the outset become hopelessly superficial as the work progresses.

Accountability systems are based on the premise that, with the proper system of incentives and sanctions, people in schools will somehow learn to do the work more effectively, and that they will learn to do it in a progressively more powerful way over time. Yet accountability systems as they are currently constructed do little or nothing to support the learning that is critical to their success. This learning is assumed to be part of the capacity that is borrowed when one level of government uses policy to make another level do something it might not otherwise do.

There are some tentative lessons from the work of improving schools and school systems that help us to understand the role of learning in accountability. The most basic of these, which sounds obvious, but is hardly ever acknowledged in the design or implementation of accountability systems, is that learning takes time. Accountability systems operate in “policy time,” constructed around the regularities of political authorization cycles, electoral cycles, budget cycles, etc. Learning in schools operates in what might be called “practice time.” Even when people are highly motivated to find solutions to problems of instructional practice—which often they are not—it takes time to get those solutions translated into instructional and managerial practice, it takes time to get those practices introduced into classrooms and schools, it takes time for the practices to result in differences in student performance, and it takes time to discover the limits of those practices and to invent the next level of practice. With practice, schools and schools systems become more efficient at these
processes, and they learn to do work that used to take them an entire school year in a few weeks or months. It is important to acknowledge, however, that policy time and practice time have no necessary relationship to each other. Solutions that are crafted in policy time—like the Annual Yearly Progress Requirement of No Child Left Behind—bear no relationship to the actual time it actually takes schools and school systems to develop, introduce, nurture, test, and redevelop new practices.

A common complaint by policy operatives and senior administrators toward school people is that they lack of a “sense of urgency” about the need to improve student performance. To be sure, teachers and principals vary considerably in their sense of urgency around the work. But what policy operatives and senior administrators usually fail to understand is that they have no basis for judging how long the work will take because there is no empirical basis for the regulatory requirements that accountability systems contain. You can’t say that things aren’t moving fast enough when you have no empirically-based standard for judging how fast they can move. For this reason, people who do share a sense of urgency about the work, and who are actually moving at a rate close to the maximum feasible, often feel that policy operative and senior administrators live in a parallel universe with no understanding whatsoever of the work.

The second lesson about learning in accountability systems is that improvement in performance is never constant and linear. Improvements in performance often—one might say usually—lag improvements in capacity and quality; that is, one sees significant changes in classroom practice well before one sees their results in student test scores. It takes time for new practices to become seated in the culture and organization of schools, and it takes time for those practices to develop to a level that can be seen in student performance. As schools improve, they pass through various levels in the development of their internal capacity to make changes in instructional practice; they get better at the improvement process as they improve. Many schools—indeed, most—from my experience get “stuck” at levels and can’t find the next the set of practices that pull them to the next level of performance. Most school systems—indeed states—manifest similar patterns. All of these are common patterns in developmental processes in the public and private sector organizations. Accountability problems develop when the system of incentives and sanctions is not congruent with the underlying
developmental processes of the organizations it attempts to regulate.

The third lesson is that learning is social. New knowledge doesn’t materialize out of nowhere and spontaneously find its way into classroom practice. Knowledge travels through social relationships. Social relationships have to be constructed intentionally, and managed intentionally, in order to produce improvements in performance over time. Teaching is complex, interactive work. We who teach learn the hard way that we are usually never accurate reporters on our own teaching practice. Learning new practices requires exposing one’s practice to external scrutiny, anchoring that practice in an external body of knowledge, and making judgments about practice based on its effects. All of these activities are social practices. They have to be learned just like any other practice. Schools have, in the past, not been places that easily accommodate such learning. So introducing new norms for adult learning requires changing the organization and culture of schools: schedules have to be changed to accommodate group work around practice, teachers and administrators have to be introduced to more systematic ways of talking to each other about their practice, new roles have to be developed to bring new knowledge into the classroom (mentor, coach, professional developer), people have to learn new skills for analyzing evidence of student learning and converting that knowledge into new practice. To the degree that you try to improve student learning without developing these practices, you put impossible pressure on individual members of the organization to all the required learning themselves. You are literally asking them to do something that they cannot do by, or for, themselves. Teachers are used to being asked to do impossible work. But this particular kind of impossible work has a mean and unforgiving nature. It says that the individual teacher is responsible for something that only the organization can do. A rational person would not agree to work in such a setting.\(^{22}\)

It is now commonplace to conceptualize learning and improvement in school systems as a kind of Chinese box

\(^{22}\) This reason, among many others, is why individual merit pay for teachers is a fatally flawed accountability mechanism. Incentives should reward behavior that results in collective learning, not individual performance.
Learning that is instrumental to improvement occurs at the individual level, at the group level within schools, at the school level, and at the system level; each level of learning is required to support the others; learning at each level is the reciprocal of learning at all other levels. So learning at the individual level—teachers and principals, for example—is instrumental to improving learning of students, but students provide the feedback that is necessary for adults to learn. Learning at the group level within schools is necessary to generate broad-scale learning at the individual level, but learning strategies at the school level are dependent on feedback from the individual and group level. Learning at the system level is necessary to induce learning at the school, group, and individual level, but learning at these levels is necessary to inform learning at the system level.

Think about a typical strategy of improvement that many school systems adopt: curriculum-driven improvement. Systems adopt ambitious new reading and mathematics curricula that are designed to produce instructional practice in classrooms that will lead to improved student performance. If the curricula are not simply restatements of what teachers are already doing, they require significant new learning on the part of teachers, and significant new learning on the part of students, both of the new content and typically of what it means to be a student. The learning required of teachers has to be supported at the group, school, and system levels. The learning of teachers is also heavily informed by what they see as students’ responses to the new content and to their teaching of it. The kinds of problems that present in the classroom—the most typical of which is that some students simply don’t respond in the predicted way to new content and pedagogy—inform the kinds of knowledge and support required at other levels if the strategy is to succeed. Strategies of improvement the school and system levels also

---

have a logic of their own: they are predicated on problems of scale that do not occur, or do not occur in the same way, at the classroom level. The choice of curriculum, professional development, and accountability strategies has to be based on judgments of cost and administrative feasibility that would not necessarily figure in classroom decisions. In the real world of school- and system-level improvement strategies there is a constant interplay of roles around these issues of knowledge and feasibility. If the purpose of an accountability system is to incite improvement, it has to be sensitive to the actual requirements of improvement processes as they occur on the ground.

Successful school improvement, then, requires very high levels of social capital: heavy investments in learning at various levels, structures and processes that link people at various levels with each other and create channels of feedback across levels, and investments in external supports and connections to sources of new knowledge. Most school- and system-level administrators have no idea what the required level of investment in social capital will be to support the level improvement that they are trying to achieve. This may be a good thing, since, if they knew, they might not undertake the task in the first place.

Improvement in performance requires progressively greater investments in social capital, scaled to the demands of increasing performance. This is not a point of view that is well-adapted to traditional forms of school organization. School systems have traditionally operated on the principle of extracting knowledge from individuals, rather than developing it. The idea was that individuals came to the organization “qualified” in one way or another for the work and the organization “used” their talents in the service of collective ends. Improvement requires a shift to a human investment view: Individuals come to the organization with a base-level of knowledge and skill, and

24 This illustrates Albert Hirschman’s famous dictum, called “the principle of the hiding hand”—the knowledge and resource requirements of complex development processes are so large that, if decision makers were aware of them, they probably would not undertake them. Development, Hirschman argues, requires a process of serial discovery, scaled to the growing capacity of individuals and institutions to do the work. Albert Hirschman, The Principle of the Hiding Hand (Washington, D.C.: Brookings Institution, 1967).
it is the job of the organization to build on that base over the course of the individual’s career to make it consistent with the demands of the work. School systems have traditionally operated on the principle that the work of teachers is isolated and autonomous; improvement requires a definition of work that is collective and cumulative over time. The structures of schools and school systems, and the human resource management systems, are all built on the old model—rewarding individual work with a standardized definition of expertise, rather than a differentiated career structure that acknowledges large differences in knowledge and skill that may not correspond to differences in experience. The form of organization required for improvement is not the one that schools and school systems have.

Getting from one form of organization to another is never graceful or easy. It requires the application of knowledge and skill over time. The first stage of this process is knowing what you don’t know. A very large proportion of school administrators have not yet reached this stage. Getting them there is another problem of building social capital.

How Systems Learn

The Boston Public Schools, like many other urban school systems, has a well-developed strategy for school improvement. The strategy was developed in a more or less cumulative way over the tenure of out-going superintendent Tom Payzant. The strategy represents most of the of principles of improvement outlined above: heavy investments in individual knowledge and skill at all levels of the system, creation of new roles and relationships organized around bring new knowledge of instructional practice into classrooms and schools, development of new career structures and new ways of developing leadership in the system, increased focus on reciprocal accountability relationships between system-level and school-level personnel, and significant infusions of new knowledge from outside the organization. Boston is a system that currently has 90 schools out of a total of 145 in various stages of sanction under No Child Left Behind. The definition of a good accountability system would be one that helps Boston get better, not one that makes it more difficult for Boston to do the work it is committed to doing.

At the school level, within the system, one can see dramatically different responses across individual schools to the district’s overall improvement strategy. A handful
of schools are exemplars—you can see every element of the improvement strategy represented in more or less exactly the form it was envisioned by the district strategy, having the predicted effect on student performance. In a significant number of schools you can see every element of the strategy faithfully represented and almost no discernible effect on student performance. In these schools the strategy is a disconnected collection of events glued onto an atomized organization with no discernible connection to classroom practice. In another collection of schools, you can see what you would expect to happen in a system predicated on organizational learning: teachers and administrators are struggling to figure out how to use the resources that come with the strategy to support their work, they are seeing changes in the quality of instruction and student performance but not always in a consistent or robust direction over time, and they are seriously engaged in trying to learn how to do the work more effectively.

This is what learning looks like in a complex system. Well-crafted system-level strategies have differential school-level effects. In general, one can see trends in performance over time that provide guidance for what is working and what isn’t. But the effects of the strategy are never uniform and never completely predictable. Large systems learn by managing diversity, not by creating uniformity. The strategy provides the base-level consistency that creates expectations for performance and assurances of support for the learning required to meet these expectations. The actual work is about responding the wide range of differential responses that schools, and the people who work in them, have to the base-level elements of the strategy. System-level administrators are constantly trying to understand and correct for the responses that their strategies generate. School-level administrators have a similar problem at a lower level: understanding and correcting for differential responses among teachers and students to a common strategy of improvement. And teachers have a similar problem in the classroom: understanding and adjusting for differential responses among students to an overall plan of instructional practice.

An effective accountability system is one that generates pressure for performance and provides support for the development of knowledge and skill required to meet those expectations. Learning in an accountability structure is most successful when it is directed at taking general expectations and adapting them to local settings;
when it creates the capacity to respond to differential effects of a common set of expectations. Sustained learning over time is what makes accountability systems work.

**Redesigning Accountability Systems**

The basic problem with No Child Left Behind is that it is a regulatory policy completely divorced from both the enforcement capacity and the basic investments in social capital required to make it work. It is a regulatory strategy predicated on the assumption of borrowing capacity at the state and local levels that doesn’t exist, or at the very least doesn’t exist in anything like the quantity or distribution that would be required for the policy to work. There are two dimensions to this problem: one is political, the other is a fundamental design flaw. I deal with the political issue in another place. 25 Suffice to say here that the current situation has been created by a major failure of political accountability in which policy makers have been allowed to engage in unchecked regulatory drift with no countervailing pressure to take responsibility for the capacity problems they have created. The design problem that flows from the political problem is how to create some reasonable equilibrium between pressure and support so that schools have to respond to external pressure for performance, but, at the same time, policy makers have to submit to the discipline of providing the capacity necessary to produce the performance.

The most basic design principle is one I have developed in other places: *the principle of reciprocity.* Simply put, the principle of reciprocity says that for every unit of performance I demand of you, I have an equal responsibility to provide you with a unit of capacity to produce that performance. Under this principle, no accountability policy would be allowed to exist at any level of government that did not explicitly calibrate demands for performance with the provision of support. So the federal government would not be able to enforce regulations preempting the design of state and local accountability systems without saying explicitly what the human investment consequences of those regulatory decisions would be and without providing financial support equal to the value of those consequences. Likewise for states. Likewise for local districts vis a vis schools. The

---

practice of borrowing capacity that doesn’t, in fact, exist is a deliberate way of shifting the cost of human investment down in the system as far as it will go until someone has to take responsibility for it. If government agencies are allowed to initiate policies without regard to the capacity requirements they impose on other levels of government, they have no reason to take these costs into account when they make policy decisions. Under these circumstances, accountability decisions become “free goods,” subject to no countervailing considerations of cost or feasibility. This is the situation we are in which we currently find ourselves.

No Child Left Behind is a regulatory regime connected to a grant-in-aid program. This raises the question of whether the grant-in-aid compensates for the capacity costs of the accountability regulations. The problem here is that the federal government is not, in the case of NCLB, only regulating the costs incurred in implementing the grant. It is regulating the entire accountability structure of public education in the country using the leverage of the grant. The right metric here for determining whether the federal government has met its responsibility for reciprocity is not the amount of the grant-in-aid. It is the capacity costs to the entire system for meeting the accountability requirements of the law. By this standard, the money that flows from NCLB is a fraction of what would be required to meet the reciprocity requirement. A federal role in accountability that is scaled to current expenditures through NCLB would require an alteration of both the substance of the law and the purpose of expenditures—a shift away from issuing regulations which it has neither the capacity to enforce nor the resources to compensate for, toward one of providing a common metric for state and local accountability systems and supporting investments in research, knowledge development and implementation for which it has a comparative advantage.

The second principle of redesign is that performance standards and sanctions under accountability systems should have to meet a test of empirical validity. This is a simple principle of regulatory policy that applies everywhere except in education. When we regulate air pollution or product safety, the standards that are set are always subject to empirical testing, and the regulatory standards are evaluated in terms of their impact on both the organizations that are subject to regulation and the public at large. The requirement for empirical testing
sets up a system of political incentives that requires the many parties to regulatory policy to present evidence to support their point of view on what the appropriate standards are and that requires a regulatory body to justify their decision based on evidence. The current AYP requirement has no basis in empirical evidence, hence it is completely divorced from any reality about the rate of improvement possible under prevailing resource constraints and the human investment costs required to meet certain performance targets. Hence, AYP produces a regulatory caseload that does not correspond to the capacity of any agency responsible for its implementation—local, state, or federal. It creates arbitrary performance targets that do not correspond to anything that anyone knows on the ground about how to improve performance; the performance standards are completely divorced from the knowledge base of the people who are responsible for meeting them. There is no public self-correcting mechanism for the performance standards affect schools in which standard-setting agencies are required to justify their decisions against any criterion other than their own intuition. AYP reminds one of the saying the during the late days of the Soviet Empire: "We pretend to work, they pretend to pay us."

Subjecting performance standards to a requirement of empirical justification would shift resources away from enforcement and toward a more systematic study of the processes of school improvement underlying accountability systems, and force a more direct discussion of the human investment requirements of meeting performance targets. There should be open study and debate about the feasibility of standards and about the conditions required to meet them.

The third principle of redesign should be an explicit reckoning with the institutional comparative advantage of governments at various levels. The federal government is currently in a position of regulating a sector for which it has no comparative advantage, and never has. There are reasons why we have never nationalized public schooling in American, and they are becoming progressively more obvious every day NCLB is in effect. The states are very weak partners in the federalization of education policy, and they are complicit in the movement toward over-regulation and under-investment in human resources. Right now those local agencies that choose to take accountability seriously are bearing the major share of the costs of both compliance and human investment, with no reciprocal recognition of what those costs are. No one is taking responsibility for
the research and development function that would be necessary to create strong, empirically-based performance standards and a human investment strategy that would make it possible to achieve them.

A well-designed accountability structure would pull the federal government out of its currently role as all-purpose enforcement officer and regulator of state and local accountability systems and put it in a role for which it has some comparative advantage. Basic issues in the design and development of instructional systems, in the technical and practical consequences of testing systems, in the development and use of professional learning systems, and in the monitoring and calibration of state and local assessment systems against international standards are obvious roles for the federal government to play. These are issues in which it is more or less inefficient to develop knowledge exclusively at the state and local levels. They are also issues that are consistent with the federal governments limited organizational capacity and its limited knowledge of the details of school improvement. States would take the major role in developing and implementing accountability systems, consistent with the principle of reciprocity, and within the context of a well-developed federal framework for what constitutes high-level standards, high-quality testing, and state-of-the-art instruction and professional development. States also have a comparative advantage in creating social capital in the form of networks that span district boundaries for the development and dissemination of effective practice. Localities would take major responsibility for the management of school improvement and the development of school improvement strategies. And schools would take the major responsibility for delivering high quality instruction, informed by performance standards and existing knowledge.

The existing accountability structure is massively over-invested in testing and enforcement and massively under-invested in capacity and social capital. The situation has come about largely as a result of unchecked regulatory drift, in which the federal government and its state agents have used the principle of borrowing capacity to consolidate and exercise regulatory authority and to push the responsibility and costs of school improvement down in the system as far as they will go without any countervailing discipline to control or compensate for the regulatory load or the human investment costs that the policy entails. This drift has created a situation in
which the human and organizational requirements for improving student learning and performance have been superceded by regulatory politics among levels of government. The result is a preoccupation among policy makers with test scores, based on no understanding whatsoever of the conditions that create performance, and a more or less constant practice of blame-shifting among levels of government to try to make the performance of schools someone else’s problem. This is an accountability system that is designed to do many things, the least of which is to improve schools.