

The Civil Rights Project



Proyecto Derechos Civiles

Why High Stakes Accountability Sounds Good But Doesn't Work—

And Why We Keep on Doing It Anyway

By

Heinrich Mintrop and Gail L. Sunderman

With Foreword By

Gary Orfield

April 2009

Copyright © 2009 The Regents of the University of California, Los Angeles

This publication should be cited as:

Mintrop, H. & Sunderman, G. L. (2009). *Why high stakes accountability sounds good but doesn't work—and why we keep on doing it anyway*. Los Angeles, CA: The Civil Rights Project/Proyecto Derechos Civiles at UCLA.

Additional copies of this report may be obtained from our website at:
<http://www.civilrightsproject.ucla.edu>

Produced with generous support from The Charles Stewart Mott Foundation.

TABLE OF CONTENTS

FOREWORD	3
Executive Summary	7
Introduction	10
The Power of Sanctions	12
Does the Sanctions System Work?	13
Does the System Produce the Expected Outcomes?.....	13
Do the Sanctions Work?	14
Is the Sanctions System Practical?	16
Simplistic Goal Setting and Misidentification of Schools	16
Insensitivity to Exclusion.....	17
Insensitivity to Special Needs.....	17
Discouraging Rigor.....	18
Turbulence	19
Strained Capacities.....	20
Coping with Impracticality	21
Is the Sanctions System Valued and Legitimate?	22
Conclusion: Why We Retain the Sanctions System and What it Costs	24
REFERENCES	27

FOREWORD

We have bet the future of federal education policy on a theory of accountability that does not work. It has been the dominant educational reform theory for decades and the No Child Left Behind Act (NCLB) is its extreme expression. It turns out, after studying research results from across the country, it does not make much sense either as a managerial or an educational strategy. It has very good intentions but often sanctions those institutions where progress is most difficult and most urgently needed rather than offer the kind of help that could really make a difference. This report, by researchers Heinrich Mintrop and Gail Sunderman, dissects the logic of high-stakes accountability policies, explores why they have failed, and concludes that the failure was not one of implementation (though that made things worse) but of the basic structure of the policy.

The Civil Rights Project has been studying the results of NCLB in six states since it was passed and has previously issued 12 reports, as well as two books and a number of articles, on its implementation and the results. Gail Sunderman has led this research. Professor Mintrop is a leading expert on the impacts of sanctions-based policies at the state as well as national level. Years ago, we showed that the standards were inconsistent and sometimes meaningless, the goals were incorrectly set and unfairly punished integrated schools and those serving English language learners and other minority groups. Our work showed that the law's assumptions about teachers and sanctions were wrong and that the sanction process was undermining the good goal of keeping experienced teachers where they were most needed. We showed early on that neither the transfer option nor the supplemental educational services provisions were working. Work we commissioned demonstrated that the dropout provisions had been gutted and that the requirements placed on the states went far beyond the limited capacities of state agencies to fulfill. We have recognized all along that the goals of more equal outcomes, good statistics on outcomes by subgroups, and a number of other provisions in the Act could be part of a good policy. The Civil Rights Project has joined many other researchers in recommending the replacement of the very narrow and arbitrary goals of test scores in two subjects with a much richer accountability scheme. When we originally raised a number of these issues we were attacked by the law's defenders, but, increasingly, the issues we raised have become part of much more broadly shared views of the NCLB experience. We believe that applying the lessons of the past eight years could produce a much more effective federal policy.

Now, as the country thinks about what to do next, it is important to focus on some fundamental design problems with the NCLB that undermine its very important goal of increasing the equity and success of American schools. The first is that it was not designed around real educational experience, nor does it utilize what research has shown about the sources of educational inequality or the possibilities and conditions necessary for reform to work. Instead, NCLB is based on the dual assumptions that children are falling behind very largely because educators don't care enough and that deadlines and strong sanctions imposed by the federal government can cure the problem so that all subgroups of children will become proficient by 2014. The second problem is that it often punishes schools that are making a positive difference for students, discouraging the staff

and undermining future prospects for the school. The third is that it has a very narrow definition of education that not only diverts attention from other vital goals but also produces a strong focus on tactics that create a semblance rather than reality of success in those limited areas. The fourth is that all schools are being required to attain goals that are impossible to attain on any broad level given what we know about both the impact of schools relative to other forces in children's lives and the distribution of talent and achievement that appears in all human populations. Finally, while the law obviously hopes schools will experience deep reform, the deadlines and yearly goals do not connect with what is actually known about the time and capacity-building required to actually turn around a school. I believe that there are good ways to correct each of these problems.

This study, commissioned by the Civil Rights Project, finds that some fundamental assumptions of the law are in error and, if continuously pursued, are very likely to do more harm than good. Since state and local educational institutions have the primary responsibility for public education—paying nine-tenths of the bills and setting most of the rules—the first requirement for federal policy should be that it does no additional harm to the public school systems. This report shows that, in that respect, NCLB falls short—not only in operation but even in its design and basic assumptions. A reasonable standard would be that a policy not weaken key institutions, not undermine support for public schools, and not try to impose impossible requirements. Historically the role of the federal government has been to encourage new initiatives, to commission research, to disseminate information and statistics, and to provide resources to the schools. The sudden decision by NCLB to define the most important subject matter, to mandate the grades that are tested, to control teacher requirements, to set detailed requirements for yearly educational gains, and to order what can be very drastic sanctions represent truly radical interventions on state and local authority. A reasonable standard for such interventions is that: 1) they are based on a solid understanding of schools and school reform; 2) they make sense to and win cooperation from the teachers, local administrators and state officials who must try to make them work; and 3) they provide the resources needed to meet the additional demands. As it now stands, NCLB does not meet any of those standards.

As a political scientist who has closely followed this law, it is obvious to me that the logic of NCLB is much more political than educational. Educators and those who spend their careers studying school reform were almost wholly excluded from the framing of NCLB and were often the target of attacks by some of its advocates. NCLB is about the politics of looking tough on educational reform. Since the 1983 Reagan Administration's *A Nation at Risk* report, the dominant style in education policy has been to *look* strong by demanding accountability, putting more focus on tests, criticizing teacher organizations, and either implicitly or explicitly blaming teachers, schools and school districts serving large numbers of poor children for inequality in educational results. One theme has been that some other means of running schools, such as charter schools, would be better because it was less public. At the same time, the law raises the pressure for schools, *by themselves*, to produce equal outcomes while other social policies bearing on the lives of poor children have been cut back. The dominant rhetoric has ignored the reality—reflected in countless studies over the past four decades—that

poverty, low parent education, poor health, and inferior segregated schools all contribute powerfully to unequal outcomes, and that those conditions can only partially be addressed inside the schools. The vast majority of a child's life is spent outside of school; students come to kindergarten with hugely unequal preparation. Ignoring the rest of children's lives and communities and expecting the schools to produce perfect equality is to expect something that is impossible and has not been achieved in any nation under any educational system.

The NCLB's strategy makes sense in political terms. Policy makers look strong by being critical, demanding and tough. They blame teachers and school districts for the striking inequalities that exist between middle class suburban, largely white schools and those serving poor and minority children. There is a tendency toward escalating demands and criticisms as the latter schools fall behind. Blaming schools and their teachers takes the pressure off political leaders (and privileged communities) to play a serious role in solving the problems of children in a society that tolerates a level of child poverty higher than any other nation of similar stature.

After the massive failure of policies adopted by the first President Bush and then President Clinton and reflected in *Goals 2000* (which were supposed to produce equal schooling outcomes in the decade leading up to 2000), the second Bush Administration and Congress adopted a much more demanding set of requirements, backed with even stronger sanctions in No Child Left Behind in 2001. It was particularly striking that this was done without the slightest evidence that it was likely to work. As was widely predicted at the time by researchers who had seriously studied school reform, that policy is a failure, substantiated by the ever-growing numbers of schools and communities officially branded as failures and sanctioned. There isn't the slightest probability that the law's goals will be met, in spite of intense pressure and widespread sanctions directed against many of the nation's most troubled schools and communities. Instead of inspiring hope and drawing our strong educators and administrators to the schools that need them the most, it gives them an incentive to leave faster so that they are not branded as part of a failed institution.

This study shows that unfounded and unattainable requirements backed by strict sanctions produce counterproductive reactions, produce massive failure, and leave states with problems they cannot solve. Needless to say, this undermines the attractiveness of teaching in the sanctioned schools and the interest of student and teachers in what often becomes a narrow drill-based curriculum.

Often the alternatives are posed as a choice between accountability NCLB-style and no accountability. This is like saying that there is no alternative between bleeding a patient and letting him die. There are, of course, better alternatives. Medicine would not dream of accepting directives from Congress that it should only pay attention to two indicators of health and that Congress could prescribe the rate of mandatory patient recovery without any reference to what the best medical research indicated was possible. This is what has happened in education. The real choice is between narrow accountability with arbitrary standards and broader accountability linked to research on what are

actually attainable goals within a prescribed period of time. Many civil rights groups and experts have signed statements advocating multiple measures which we strongly support.

Our survey of teachers in two highly impoverished school districts actually showed that teachers supported accountability but they wanted reasonable standards, materials that effectively addressed the learning challenges of students, time to work together to back the school's educational operations, and a good professional leader in their building. Federal law should foster these goals. What we need are richer forms of accountability measuring: school completion; acquisition of key skills needed for college; encouragement of a full education rather than endless drills on just two subjects; promotion of the kind of imagination and activities that attract students to learning; preparation of our students to become capable citizens in a democracy, and rewards for teachers and principals who make a clear difference in the level of success for their students.

As we go into the next round of debate it would be very valuable to go back to first principles and consider how the federal investment in education could be refocused in a more positive way. It is very clear that there were tremendous inequalities in our schools before NCLB that still exist and still need to be addressed. I think that a strong and positive federal role is an important part of the solution, but that it must be developed in collaboration with those who truly understand schools and work in them, that it must provide resources to build capacity, that it should support building more knowledge in a field where there are too few proven remedies, and that it should provide recognition and support for those schools, administrators, and teachers whose work makes a real difference for children. This study by two leading scholars in the field is a solid contribution to that effort.

Gary Orfield
Co-director, Civil Rights Project/Proyecto Derechos Civiles

Why High Stakes Accountability Sounds, Good but Does Not Work— And Why We Keep on Doing It Anyway

Executive Summary

The federal accountability system, made universal through the No Child Left Behind Act of 2002, is at its heart a quota and sanctions system. This system stipulates the progression of underperforming schools through a set of increasingly severe sanctions based on meeting performance quotas for specific demographic groups. While it includes standards, assessments, and performance targets, sanctions are the means by which the higher levels of the system put pressure on lower-levels of the system to take accountability seriously. Even though the law formulates the sanctions in the language of improvement, support, and radical renewal, the punitive core for districts and schools is apparent. When improvement efforts fail, loss of control and threat of organizational survival is at stake.

Whether this system is up to the job of achieving its goal—improving the performance of persistently underperforming schools—is an open question. Using findings from the best available research, this report examines whether an accountability system based on the imposition of sanctions is likely to succeed or fail and, if it does persist, what the consequences may be for sustaining an educationally rigorous system. The report asks three questions: (1) does the system work, that is, does it produce the expected outcomes; (2) is it practical, or can it be implemented, and (3) is it legitimate, or is it valued among those who must implement it. We conclude with a discussion of the costs of maintaining the current sanctions system.

Does the System Work?

Does the System Produce the Expected Outcomes? There is little evidence that high stakes accountability under NCLB improves student achievement. Although state accountability systems appear to be a success since state test scores continue to rise in most systems, the picture looks far less positive when one looks at the National Assessment of Educational Progress (NAEP). When NAEP scores are used, gains appear to be much lower. Substantial variation among states exists, and few states have narrowed the achievement gap among racial and socioeconomic subgroups while simultaneously improving overall performance. Given the large discrepancies between NAEP and state assessments results, it is not quite clear what the state tests measure. By all indication, state accountability systems, with their own pressures and sanctions, are successful at focusing schools' and districts' attention on state assessments.

Do the Sanctions Work? There is also a lack of evidence that the sanctions themselves have been successful as an effective and universal treatment for raising achievement levels at low-performing schools. Neither the transfer option nor the supplemental educational service provisions have been widely embraced by parents or districts. Whether or not the transfer option produces improvements in school

performance is a moot point, since the percentage of students taking advantage of this option (about 1% of eligible students) is so low. The response to supplemental educational services has also been low (14% of eligible students). Third party evaluations of these services find small, if any statistically significant effects of the program on improving student achievement. The corrective actions and restructuring options, such as reconstitution, charter school conversion or take-over by education management organizations (EMOs), may work in some limited situations but are not effective across the board. Among the variety of corrective actions and restructuring strategies already tried, none stick out as universally effective or robust enough to overcome the power of the specific conditions of local implementation contexts.

Is the Sanctions System Practical?

If the NCLB system was practical, it would do the following: identify schools in need of improvement and restructuring with high accuracy; appropriately direct schools to pay attention to students most in need of help; produce an intervention burden for states and districts commensurate with capacities to provide new impetus, ideas, resources, and personnel, and lastly, through the imposition of sanctions, create momentum for deliberate and well-articulated improvement processes for schools and districts stuck in low performance. NCLB fails these practical criteria.

In state systems with at least moderately high performance demands, NCLB has labeled high numbers of schools as failing, which has far surpassed district and state capacity to intervene. But it is not even clear if the bulk of these schools are in fact correctly classified. Most notably, the system offers no practical answers to address the full spectrum of student performance and learning needs, particularly for students far below proficient, special needs students, and marginally performing students. Moreover, it does not speak to the predicament of low-capacity schools and districts. While it may appear that the sanctions system has succeeded in fermenting a climate of reform, such ferment, in many instances, is more likely to result in unproductive turbulence than sustained school improvement.

Is the Sanctions System Legitimate?

Despite an almost twenty-year period in some states, accountability systems, and particularly NCLB, continue to encounter serious legitimacy and acceptability problems among the groups that they are designed to target: teachers, principals, and administrators in low performing schools and districts. In general, while standards, assessments of performance, and consequences for low performance are widely accepted ideas *in general*, research suggests that attitudes about high stakes accountability systems are more negative. Accountability systems designed around sanctions violate core professional norms of educators and produce widespread frustration and de-moralization among those charged with carrying out school improvement efforts. Accountability goals are often not seen as realistic, and the sanctions are considered to be misguided and not very useful for improving schools. In efforts to improve test scores, teachers widely

report that they must compromise standards of good teaching in order to meet accountability goals.

What are the Costs of Maintaining a Sanctions System?

The combination of uncertain effects, loose connections to the broader educational values and norms of educators, and the difficulties or impossibilities of carrying out the system day-to-day makes the sanctions system a prime candidate for declaring it a failing system. But there is a way to maintain the system, although this approach produces high educational costs. As long as states maintain low-rigor systems that concentrate on basic skills, and the more lenient options for school improvement or restructuring are chosen, the system can persist with relative ease. NCLB “works” when systems place low demands on the cognitive complexity of learning tasks and, subsequently, on teacher capacity building. State accountability systems that operate within a basic skills framework and with low test rigor tend to produce lower numbers of failing schools. Because such systems tackle school improvement goals that are fairly light, affordable, and manageable, they are more practical within the NCLB framework. Systems that are more ambitious produce an intervention burden that makes them unworkable.

Improvement strategies that may be sufficient to reach lower level goals are not sufficient to reach higher order goals. Once educators have operated within the confines of a system oriented around test-driven basic skills remediation, strategies for teaching and school improvement cannot simply be switched to higher level skills, with the result that whole state systems get stuck in low level intellectual work.

A redesign of the federal accountability system should start from four principles:

- The system should reflect the complexity of the task by allowing multiple measures, more flexibility and local options.
- Ambitious goals require ambitious capacity-building within schools and districts, and in all likelihood beyond.
- More comprehensive investments in student welfare that link education with health, job development, and community building, as well as redistributive investments to attract and keep top-flight professionals in poorly performing schools, are needed.
- Overreliance on sanctions can be reduced when policies aim to develop a partnership between government, teachers, and parents, and motivate changes by adhering to the professional values and standards of educators.

The report contends that fifteen years of state and federal sanctions-driven accountability has yielded relatively little. It is time to try a new approach that replaces a system based on mandates and legal administrative enforcement with one that emphasizes the professionalism of educators and the active involvement of communities.

Introduction

Sanctions are a fact of life. When children do not do their homework, they lose points. When employees do not come to work, their pay gets docked. When shopkeepers do not wash their display windows, customers will shun their wares. Sanctions make intuitive sense: we want people to do the right thing, and we feel it is only fair when they bear the consequences for lack of performance. Industrial psychologists and organizational sociologists have shown how rewards and sanctions function at the core of work performance (Cooper & Robertson, 1986; Latham & Pinder, 2005; Lawler, 1973). But how rewards and sanctions play out has differed quite remarkably over time (Sennett, 2006). ‘Organization man,’ working within stable private or public bureaucracies, could look forward to climbing the career ladder given adequate work effort, loyalty to the organization, and seniority. Being passed over for a promotion at the normative time may have been financially disadvantageous and socially shameful, but one did not lose one’s place in the organization. Nowadays, the stakes are higher. Those who are nimble and flexible with up-to-date skills can reap enormous rewards while those who cannot keep up in the competition are apt to lose job, status, and livelihood. Today’s high performance work organizations are of this decidedly more high stakes type. While they dangle high rewards for some, they foster a climate of punitive uncertainty for others, particularly those who find themselves working in struggling or losing industries.

Not unlike large corporations in the business world, schools in the past were organized as stable hierarchies, though with rather flat career ladders that rested on faculty members’ solidarity and attachment to a given school or school district. Tenure and seniority increased longevity in the job and decreased uncertainty, envy, competition, and turnover. The “apprenticeship of observation” (Lortie, 1977) laid the groundwork for socializing new teachers into established performance standards. Group solidarity among teachers shored up collective expectations of average or middling effort; poor performance of a few ‘bad apples’ was tolerated and high performance was largely ignored.

A decade or so after corporate restructuring, the new high-stakes work organization arrived in the public school system with the advent of accountability systems, first introduced in a few states in the early nineties, and then made universal through the federal No Child Left Behind (NCLB) system. Interestingly, after a few experiments with the ‘rewards’ aspect of high stakes accountability, most notably in Kentucky and California, school accountability systems gravitated to the ‘sanctions’ side (Mintrop & Trujillo, 2005). With the additional prod of NCLB, state accountability systems for the most part narrowed their mission. Increasingly bypassing schools that tended to enroll more middle class students, state accountability systems instead targeted, with ‘laser-sharp focus,’ the lowest performing schools with the goal of closing the achievement gap. As they became more equity-oriented, they also became more punitive.

The prevalence of incentives and sanctions is tied to a new centralism in goal setting and goal monitoring made possible by new information and data warehousing technologies (Sennet, 2006). It has now become practical for a central planning agency,

be it private or public, to set targets based on a small set of quantitative performance indicators, monitor whether large numbers of relatively small performance units reach these targets and surgically order sanctions for underperforming units. This new performance system increases central control by top management, freezing out the mediating functions of middle layers of the organization. In the political realm, the system increases the possibility of a small group of centrally positioned elites to steer a whole system, whether these be efficiency-oriented politicians or equity-oriented civil rights leaders. But in contrast to private corporations, in education, accountability is grafted onto, and runs up against, the traditional loose coupling of the educational system, political contestation at various layers of the system, and the enduring unionization of teachers. In private industry, the new performance systems could be bolstered by a good dose of coercion. In the educational system, the coercive power of the center is greatly reduced, making the success of the system much more incumbent on legitimacy.

The current accountability system in education consists of standards for subject matter content and skills, standardized tests as performance indicators, and targets and quotas for measuring performance and underperformance. Sanctions are the means by which higher levels of the system put pressure on lower-level performance units—schools, districts, and states—to take central demands seriously. We are talking here of a system in the sense that a central authority uses legal, administrative, and institutional arrangements to maintain control over subordinate layers of the system in the service of an effective delivery of public goods (Aucoin & Heintzman, 2000). Accountability systems are one such type of system that compels schools and districts to abide by centrally set performance norms.

At its heart, the federal accountability system is a quota and sanctions system. The substance of academic content, testing rigor, regulation of inputs (with the exception of the requirement for highly qualified teachers), specific method of school or program improvement and/or restructuring: these elements are left for states to decide, as long as they all exist formally. What is strictly stipulated, however, is the staged progression of underperforming units through a set of increasingly severe sanctions based on meeting performance quotas for specific demographic groups. This begins with the identification and publication of “school improvement” status (a kind of public shaming with potentially far-reaching market consequences), then leads to the mounting loss of organizational autonomy through required external intervention and service contracting, and finally termination through re-organization or take-over of the organization. Even though the law formulates the stages of sanctions utilizing the language of “improvement, support, and radical renewal,” the punitive core for districts and schools is apparent. When improvement efforts fail, loss of control and threat of organizational survival is at stake.

Raising the overall achievement of a whole national educational system and closing the achievement gap is obviously an enormously complex problem. NCLB is the simple policy answer to that complex problem that currently holds sway. If indeed a combination of quotas and sanctions connected to new data processing and warehousing technologies could do the job, we would be the first to support such a system given the

dire situation of many poor students and students of color in the nation's schools. But we argue that, at this juncture, the cracks of failure in the federal sanctions system have appeared, and the simplicity of the remedy has become increasingly overshadowed by the complexity of school and district improvement. But since there is an "easy" way to make the system work, it may be retained anyway.

The Power of Sanctions

"Accountability is here to stay" was a slogan often heard in the early nineties to persuade disbelieving teachers to make the necessary adjustments. State accountability systems have shown staying power—certainly beyond what critics imagined. By all indication, they have proven to be powerful instruments in reshaping how schools go about their business, especially in schools for poor and minority children (Au, 2007; Herman, 2004). Accountability systems divide into two main components: 1) guidance in the form of standards and test data that orient instruction and inform performance, 2) and firm performance targets, pressures and sanctions that make the system compelling.

By their very nature, pressures and sanctions should be perceived as more negative than standards and tests, the former being more controlling, the latter more informative (Frey, 1997). Sanctions are penalties for non-compliance with authoritative regulations or powerful demands. They may inflict loss of benefits, prestige or status on individuals or collectives, and trigger attendant feelings of displeasure, shame or fear (Posner & Rasmusen, 1999). In the extreme, they threaten freedom or survival. Sanctions systems can be quite costly as the outlays for the criminal justice system in the US demonstrate. Costs are reduced when the actual imposition of sanctions is the exception, the threat of sanctions is sufficient to compel the desired behavior, or the expected behavior occurs largely voluntarily. Sanctions are credible when they are properly targeted on those actors who are responsible for expected behaviors, when they cause clearly unwanted discomfort, and when they can be enforced. Voluntary compliance is more likely when the expected behaviors are valued, because they "work," i.e. lead to expected outcomes, or correspond to personal values, dispositions, or ideologies, and when actors have the requisite capacities to fulfill expectations (Coleman, 1990; Lawler, 1973; Shamir 1990, 1991).

Conversely, sanctions systems are likely to fail when they produce ambiguous or uncertain outcomes so that their effect is in doubt. Sanctions are likely to fail when designed with less-than-credible threats, aimed at diffuse target actors, beset with visible enforcement challenges, and requiring behaviors for which there is not sufficient capacity. Under these conditions they become impractical. Legitimacy will be compromised when sanctions only tenuously correspond to target actors' values and norms, befall actors who do not feel at fault, or instill doom as opposed to hope or expectation of success. Sanctions lacking practicality and legitimacy may induce defensiveness (Argyris, 1990; Staw et al., 1981) that prevents organizational actors from learning and trying out new solutions. The systems nevertheless may still be maintained (Meyer & Zucker, 1989) as long as secondary benefits can be derived from them, such as symbolic or ideological satisfaction, economic gain, or simply because of the lack of

realistic alternatives that address the problem as forcefully as the sanctions system, on the face of it, promises to do.

We argue in this paper that we have arrived at a fork in the road:

- the federal sanctions system has developed the adverse conditions enumerated above, in which case its failure becomes predictable;
- it may be fixed or made to work, in which case it survives, but its survival becomes educationally undesirable, or
- it hangs on despite its failure, propped up by a coalition of secondary beneficiaries with political power.

In the following sections we elaborate our arguments by asking three main questions: (1) does the system work, that is, does it produce the intended results, (2) is it practical, that is, can it be implemented, and (3) is it legitimate.

Does the Sanctions System Work?

This question divides into two aspects: does the system as a whole produce the expected outcomes, and do the actual sanctions do the job. The question is not whether sanctions-based accountability has had any effect. Some effects can be assumed. The question is whether the gains justify the costs—the negativity of the sanctions system, the disruption created by restructuring, and the loss of control of local actors over their affairs. This is doubtful.

Does the System Produce the Expected Outcomes?

To date, scholars argue whether the law has had the powerful impact on schools for which it was designed. NCLB was implemented amid controversies over whether external, test-based accountability improves or hinders student achievement (Amrein & Berliner, 2002; Carnoy & Loeb, 2002; Haney, 2000; Hanushek & Raymond, 2004; Lee & Wong, 2004; Raymond & Hanushek, 2003). Research on the effects of high-stakes testing and accountability policies adopted prior to NCLB was mixed, inconclusive, and often contradictory, and there was little information on the mechanism through which accountability policies may affect student achievement or the achievement gap (Lee, 2007). These studies found larger or smaller effects depending to a large degree on the statistical methodologies they employed and the way they imputed students' exclusion from testing.

Whether or not high stakes accountability under NCLB improves student achievement also depends in part on the metric used. Most state accountability systems appear to be a success, given that in most systems test scores continue to rise. In fact, analyses that use state assessment results show increases in overall average scores (Center for Education Policy, 2008b; Neal & Schanzenbach, 2008; The Education Trust, 2004, 2005). The picture looks far less positive when one looks at the National Assessment of Educational Progress (NAEP), the only cross-state metric we have. The

idea behind comparing state test results to NAEP is that if gains on high-stakes state tests represent real gains in achievement, they should generalize to low-stakes tests, such as NAEP (Koretz, 2008; Lee, 2007). When NAEP scores are used, gains appear to be much lower (Fuller, Gesicki, Kang, & Wright, 2006; Fuller, Wright, Gesicki, & Kang, 2007; Lee, 2006, 2007). The picture for elementary grades seems better than for secondary grades and better for math than for reading—reading scores tend to remain flat, while grade 4 math scores showed some improvement. Nonetheless, there are substantial variations among states, and few states have narrowed the achievement gap among racial and socioeconomic subgroups while simultaneously improving overall performance (Lee, 2007).

While NAEP scores have risen since NCLB, it is difficult to attribute gains to NCLB simply because the scores represent trends that began prior to NCLB and do not reflect any significant acceleration in the pace of improvement after NCLB passage. Lee (2007) examined long term trends in national average math and reading scores and in the achievement gap between 1971 and 2004. He found small or moderate improvements in both reading and math, but no indication that the improvements in achievement were related to any educational reform policies (Nation at Risk in 1983, Goals 2000, Improving America Schools Act in 1994, and NCLB in 2001). The trend lines were linear and there were no significant changes in the performance trajectories over the entire 33 year history of NAEP. Long term trends in reducing the racial and socioeconomic achievement gap showed a curvilinear pattern, with reductions in achievement gaps in the 1970s and 1980s and an increase in the 1980s and 1990s. Since NCLB passage, the gap has not narrowed significantly.

Given the large discrepancies between NAEP and state assessments results, it is unclear what the state tests measure. By all indication, state accountability systems with their own pressures and sanctions are successful at focusing schools' and districts' attention on state assessments. The literature is replete with accounts of schools' attempts to streamline or "align" teaching with state demands. But whether students actually learn more and are given a better education in higher performing schools is doubtful (Mintrop & Trujillo, 2007). It has become increasingly apparent that teachers in low-achieving schools, who must generate larger gains than high-achieving schools, have strong incentives to adopt practices that inflate test scores (Koretz, 2008).

Do the Sanctions Work?

The record as to whether the sanctions themselves work is equally ambiguous. Data on school transfers and supplemental educational services, the two high-profile interventions in the early improvement stages, suggest that parents for the most part ignore the new options, or districts are either unwilling or unable to accommodate parents' wishes. Most districts required to offer transfers lack alternatives that are better than the schools students attempt to transfer from, making it difficult for districts to implement the policy (Kim & Sunderman, 2004). Whether or not the transfer option produces improvements in school performance is a moot point since the percentage of students taking advantage of this option (about 1% of eligible students) is too statistically

low to reliably attribute school improvement effects to it. More students have taken advantage of supplemental educational services (14% of eligible students in 2005-06) (Hoff, 2007), but the response has not been overwhelming and third party evaluations of these services are finding small, if any statistically significant effects of the program on improving student achievement (Chicago Public Schools, 2007; Heinrich, Meyer, & Whitten, 2008; Heistad, 2006; Potter et al., 2007; Rickles, Barnhart, & Gualpa, 2008; Zimmer, Gill, Razquin, Booker, & Lockwood, 2007).

Corrective action and restructuring options under NCLB, such as reconstitution, charter school conversion or take-over by education management organizations (EMOs), sometimes work and sometimes do not (see Mintrop and Trujillo, 2006, for more detail.) In Maryland, for example, some local reconstitutions actually exacerbated schools' capacity problems, reduced schools' social stability, and did not lead to the hoped for improvements, although a few schools benefited from a fresh start (Malen et al., 2002). Results from reconstitutions in Chicago (Hess, 2003) and in New York's "Schools Under Registration Review" (SURR) program were inconclusive as well (Brady, 2003; NYSED, 2003). School take-over by educational management organizations (EMO) has worked in some cases, but not in others, as research from Baltimore and Philadelphia suggests (Blanc, 2003; Bracey, 2002; Molnar, 2005; Travers 2003; Useem, 2005; Saltman, 2005). State takeovers of entire districts have also produced uneven outcomes. Financial management is often cited as the most promising area for potential success of district take-over by states (Garland, 2003). However, equally dramatic *academic* success has been much harder to achieve (Ziebarth, 2002; Education Commission of the States, 2004). While the research base on charter schools is expanding, little is known about charter school conversion as a means of corrective action and school redesign (Bulkley & Wohlstetter, 2003). Available data seem to suggest that converting district-administered schools into charter schools has had uneven results (Brown Center, 2003; Gill et al. 2007). Charter schools also tend to show up on states' lists of failing schools in larger proportions than regular public schools (Brown Center, 2003).

Thus, among the variety of corrective action strategies that have been tried, none stick out as universally effective or robust enough to overcome the power of local context. Competence of local provider personnel, intervention designs, political power of local actors in the system, and district and site organizational capacity to absorb the strategies all strongly influence how a particular strategy will turn out.

The avoidance by states or districts to embrace the strong tools provided to them by NCLB (Center for Education Policy, 2007, 2008c) may be telling in this context. The school restructuring options have not been widely adopted, with schools and districts preferring more traditional school reform strategies, such as attending to how instructional time is used, hiring coaches to improve instruction, increasing staff collaboration, or providing school-based tutoring. In a survey study of 340 districts conducted by the Center for Education Policy (2007), district officials cited their own strategies as more effective in improving student performance than the more radical corrective measures stipulated by the federal sanctions system. Finally, if the NCLB sanctions "worked," then we might expect schools to move out of improvement status in

large numbers, but in many states, this is not happening (Center for Education Policy, 2008a; Owens & Sunderman, 2006).

In sum, if there were clear-cut evidence of convincing student-learning gains, the debate about the current sanctions system would end, even if a causal connection between specific sanctions and test scores could not be found. But evidence of improved student learning is ambiguous and the effectiveness of the prescribed sanctions for school improvement is mixed. It is safe to say that as of now a universally effective treatment for low-performing schools in the corrective action and restructuring stages has not materialized. Thus, we are left with great uncertainties as to NCLB's effect on student achievement—uncertainties that undermine the justification for costs associated with the law. For those committed to the sanctions system, there may be grounds for a more positive answer: state test scores keep rising and NAEP at least moves in the right direction. This kind of evidence may quell their doubts about the sanctions system, but it does not silence those with second thoughts.

Is the Sanctions System Practical?

In asking this question, we are less concerned with effects and more focused on the processes of productive school improvement that NCLB is designed to leverage. If the system was practical, it would: identify schools in need of improvement and restructuring with high accuracy; appropriately direct schools to pay attention to students most in need of help; produce an intervention burden for states and districts commensurate with capacities to provide new impetus, ideas, resources, and personnel, and lastly, through the staging of sanctions, create momentum for deliberate and well articulated improvement processes with ever higher intervention intensity in schools and districts stuck in low performance. NCLB, in our view, fails all these practicality criteria.

Simplistic Goal Setting and Misidentification of Schools

Evidence on federal goal setting in terms of adequate yearly progress towards proficiency has shown much movement around the cut-off points, but is unable to register growth in the below-proficiency band, where many high-poverty schools operate (Neal & Schanzenbach, 2008). Other research shows that NCLB is insensitive to the exclusion of students (Balfanz & Legters, 2008; Amrein and Berliner 2002; Haney 2008), creates an undue burden for schools with highly heterogeneous student populations, particularly schools with large numbers of special need students and many sub-groups (Sunderman, Kim, & Orfield, 2005), and results in an untenable number of failing and misidentified schools in all but the most lenient state systems.

Adequate yearly progress (AYP) is the measure used to hold schools and districts accountable. Schools that make AYP are assumed to be functioning well. As it turns out, AYP is not very good at differentiating schools that are making progress from those that are not. There are a number of technical reasons for this, most notably the fact that AYP compares the current proficiency status of a school or district to a fixed annual target. According to this metric, schools report the percentage of students who are performing at

or above the proficiency target for a given year. Thus, AYP is not a measure that captures improvement, or gains in student achievement, from one year to the next (Linn, 2008). Because students in schools identified for improvement for the most part begin with lower average test scores, they can continue to make substantial improvements while failing to reach the fixed AYP performance targets. As a result, overall student achievement gains are often similar in schools identified for improvement and schools meeting federal AYP goals. For example, an analysis of data from Virginia found similar levels of improvement in proficiency for both types of schools (Kim & Sunderman, 2005). If AYP was one *among many* school quality indicators used to gauge school and district performance, these inaccuracies might be tolerable, but since it is the only gauge used, it is not. As the sole authoritative indicator that triggers sanctions, AYP creates powerful realities for schools and districts on the ground. Irrationally, many schools may be forced to own their low performance label when they are in fact fairly healthy and still making progress.

Insensitivity to Exclusion

Given that extra-school factors have a large effect on variation in student achievement and school performance (Scheerens and Bosker, 1997), schools can substantially improve their chances in the system, without improvement efforts, by manipulating the composition of students taking the test. NCLB was designed to identify schools that consistently fail for low-income and minority students. At the high school level, it was intended to identify schools where students were not performing at the proficient level of academic achievement and were not graduating with a regular high school diploma in the standard number of years. Yet, among the nation's lowest performing high schools, 41% actually made AYP for both achievement and graduation (Balfanz & Legters, 2008). These are schools where the freshman cohort shrinks 40% by the senior year and a majority of students fail to graduate (Balfanz & Legters, 2004, 2008). Retention of students prior to key testing grades, reassigning weak learners to non-accountable categories, or pushing students out of school (Haney, 2008; Booher-Jennings 2005; Anagnostopoulos, 2003) are examples of exclusion coping strategies. Schools are thus "encouraged" by an insensitive measurement system that rewards and penalizes by merely registering proportions of proficient students.

Insensitivity to Special Needs

The disaggregated reporting of test score data by subgroups has brought greater transparency to the performance of minority and disadvantaged students, but this seemingly simple and straight-forward mechanism also places a very high burden on schools and districts that serve a diverse student population. By holding schools accountable for subgroups of students—English language learners, students with disabilities, low-income students, and members of all racial/ethnic groups—NCLB promised to close the achievement gap. Schools must reach performance targets in reading and math for all students and for each subgroup of students, and they must assess at least 95% of their eligible students. Under this scheme, at a minimum, there are at least five targets and could be as many as thirty-seven depending on the number of sub-

groups a school has (Marion et al., 2002), increasing the likelihood of failure in proportion to the number of sub-groups. Thus, the NCLB approach makes it considerably harder for schools with diverse student populations to make AYP, simply because they have more targets to make (Kim & Sunderman, 2005; Linn, 2005; Novak, 2003).

Meeting sub-group targets for students with limited English proficiency (LEP) or with learning disabilities has turned out to be a particularly vexing problem for schools and districts. Researchers have identified a number of challenges to implementing the NCLB requirements for LEP students, including the instability of the LEP subgroup, the failure of standardized test scores to accurately reflect what LEP students understand, and the lack of proven accommodations that would make these scores more reliable, among others (Abedi, 2004; Batt, Kim, & Sunderman, 2005; Coltrane, 2002; Kieffer, Lesaux, & Snow, 2008). State and local education officials also questioned the fairness of the provisions. Students who achieved English proficiency are generally moved out of the subgroup, while new students with very low levels of English proficiency are continually added to the subgroup, greatly diminishing the chances that schools serving large numbers of LEP students will be able to improve the performance of this subgroup and make AYP. In addition, states were finding that schools reporting an LEP subgroup were more likely to be identified as needing improvement than those without this subgroup, an issue also found with the subgroup of students with disabilities as well (Batt et al., 2005; Sunderman, Kim, & Orfield, 2005). It is not surprising that a system inclined to capture performance with highly standardized and simple measurement tools, and averse to exceptions from uniform proficiency goals, would come under enormous strains in dealing with special needs students. But the remedy is not straightforward. In the logic of the system's incentives, any group excluded from, or relieved of, accountability provisions runs the risk of being shortchanged as it is in schools' and districts' self-interest to ration their services according to the most pressing accountability demands.

Discouraging Rigor

The practicality of the sanctions system is fundamentally a function of the testing rigor of state accountability systems. Michael Fullan's idea of goal horizons (Fullan, 2003), drawn from the evaluation of national reform efforts in English schools, is useful here. He distinguishes between two goal horizons. In Horizon 1, systems concentrate on basic skills. Alignment of curricula to state tests, program prescription, tightening up basic organizational effectiveness, and all sorts of short-term strategies (such as targeting students at statistical cut-off points, teaching to the test, standard-specific remediation, pep rallies at testing time, etc.) can have initial significant effects on school improvement as measured by test scores. But the gains do not seem to persist (Fullan 2003; Mintrop 2004). Within Horizon 2, systems tackle more advanced goals, such as reasoning, higher order thinking, complex problem solving, and students' engagement in learning beyond recall and repetitive practice. While achieving basic proficiency for most students (Horizon 1) is a great challenge and worthy goal for many schools, Horizon 2 requires both students and teachers to be more deeply involved in learning. Improvement

strategies that may be sufficient to reach Horizon 1 goals do not necessarily suffice to reach Horizon 2's higher order goals.

State accountability systems that operate within Horizon 1 seem to be more practical within the present architecture of NCLB. Operating with fairly low test rigor pegged to presently available state and teacher capacity, these systems produce low numbers of failing schools and capacity-building needs that are fairly light, affordable and manageable (Paterson & Hess, 2006). Pressure may be sufficient to prod schools towards reaching the system's modest proficiency goals. Thus, in low-demand or low-rigor systems, a mere sanctions system of the NCLB type would work. In contrast, systems that are more ambitious in their performance demands produce an intervention burden that seems to make the system *unworkable*. To illustrate, data from the 2003-04 school year (Mintrop 2007; Mintrop and Trujillo 2006) show that states with high testing rigor in which the definition of proficiency is relatively close to NAEP, such as California, had up to a quarter of their schools in federal school improvement. States with large gaps between state-defined and NAEP-defined proficiency, such as Texas (where the gap is between 50 and 60 points), had a much smaller intervention burden. Only about 5 percent of the schools in Texas were identified as needing improvement. Kentucky is a state with medium testing rigor and correspondingly a medium intervention burden, with about 10 percent of total number of schools in 2003-04 needing improvement.

States that adopted accountability systems prior to NCLB (known as first generation accountability states) tended to deal with these high intervention burdens by scaling back their programs. After an initial period when some states classified large numbers of schools as needing improvement, most settled on an intervention burden of no more than 2 to 4 percent of total number of schools in the state. But this winnowing down took place prior to NCLB, when states had flexibility in how many schools they wanted to sanction as low-performing. It is no longer available under the NCLB system's firm staging of sanctions and increasing proficiency targets. Thus, if states were to adopt definitions of proficiency close to NAEP, as California did, and as some advocates have demanded of all states, the result would be a staggering number of schools in need of school improvement for which enormous intervention capacity would have to be provided. Indeed, by 2008, 48% of schools and 61% of districts in California were in improvement status (Asimov, 2008). Short of that, states have the option to keep their school accountability systems well within Horizon 1, or else risk impracticality.

Turbulence

The rigid staging of sanctions promises to result in unproductive turbulence, rather than in forceful and deliberate improvement. Schools that are persistently unable to meet AYP are not virgin reform territory for the most part. Many persistently low-performing schools are not stable in their stagnation, but volatile and continuously reconstituting in an unplanned way. Teacher and administrator turnover is often high, external consultants plentiful and ever-changing, and district intervention intensified (Mintrop, 2004; Neild & Spiridakis, 2002; Neild et al., 2003). Turnover in district

administration is said to be on the increase as well (Buchanan, 2006), not to mention the well-known revolving door at the helm of urban districts. Increasingly, urban schools are staffed by incoming cohorts of teachers and administrators who rapidly turn over (Johnson, 2007). They also tend not to conform to the (stereo) typical tenured and inert senior teacher supposedly in need of some “shaking up.” For these new cohorts, it is less a question of whether their feet will be held to the fire but more if they will get them wet in the first place. None of the reorganization measures commonly associated with federal (and state) sanctions addresses these problems.

In all likelihood, many low-performing schools, unable to meet federal AYP, will have previously been subjected to substantial local reform measures. Districts that anticipate state action and carry out local school restructuring often move principals and staff, conduct inspections, and mandate programs before a school appears on the state or federal radar screen. When that happens, schools may have to repeat improvement stages or cycles once they enter federal or state corrective action, and they may have to adopt, yet again, corrective action and restructuring strategies with uncertain and contingent prospects for improvement. Something of this nature is bound to happen in places like Philadelphia, where a fairly large number of the lowest performing schools will make their journey through the NCLB stages as already redesigned schools (Travers, 2003b). This has already happened in districts with a long history of reconstitution, such as San Francisco. As was pointed out above, charter schools tend to show up on states’ failing schools lists in larger proportions than regular public schools. For these schools as well, fundamental redesign happened prior to the introduction of the federal sanctions system.

In other words, NCLB interventions will increasingly look like *déjà vu* to affected schools, with more hoops to jump through, unless states design intervention approaches that are truly different from “all the other things” a school has already tried. Such approaches need to decrease turbulence, rather than add to it. But the rigid staging of the federal sanctions system makes designing measures appropriate to the developmental needs of a given school or district so much more difficult.

Strained Capacities

The impracticality of the NCLB system is greatly exacerbated when schools, districts, and state departments of education do not have the capacity to meet the performance demands of the accountability system. This, of course, is a more likely situation in state systems with more rigorous and cognitively complex performance demands. Capacity building is not only about direct material support for lagging schools and districts, but also about the creation of a school improvement infrastructure that is adequate to the urgency and demands of the system.

High stakes accountability systems seem to intensify a two-tier structure of high- and low-capacity schools and districts. Research has found that high-capacity schools often already possess the capacity and resources needed to perform at high levels. They are thus able to use the additional impetus and guidance from the accountability system to respond as expected—that is, to improve instruction and curriculum (Diamond &

Spillane, 2004; Elmore, 2004; Sunderman, 2001). They are therefore more likely than low-capacity schools to avoid the negative repercussions of the sanctions system.

Many poorly performing schools lack the resources and capacity to respond to sanctions on their own, and in ways that will improve curriculum and instruction (Elmore, 2004). Intensifying pressure through sanctions will not result in improvements, but further fragmentation and deterioration (Mintrop 2004). Low-capacity schools are predestined to bank on short-term strategies that require little change in capacity (Sunderman, Tracey, Kim, & Orfield, 2004). Common strategies include test preparation activities, content alignment, and concentration on tested subjects, benchmark grades, and students near proficiency. In some low-performing schools, this can amount to a parallel test-remediation curriculum that is different from the regular curriculum taught in less-pressured schools, with the result that students are excluded from intellectually challenging content and learning (Diamond & Spillane, 2004; Sunderman, 2001; Valenzuela, 2005). In low-rigor, low-demand accountability systems, these strategies might actually work to keep a school from facing corrective action, but it is unlikely that they suffice in more rigorous accountability systems.

When mere threat of sanctions is not strong enough for schools and districts to meet AYP goals, or when it has become a detriment to schools' chances for improvement, a support infrastructure is needed. NCLB relies on state education agencies to play a crucial role in implementing the federal mandates, but provides relatively modest resources to help them do so. Under NCLB, states are required to develop testing and accountability systems that in many instances go beyond what they had in place previously. They must collect and publish data on student achievement that includes disaggregation by subgroup categories and teacher quality, and is more extensive than previous data requirements. Even more importantly, states are required to help schools and districts improve under NCLB, a role that traditionally has not been a state function. State education agencies are relatively small agencies that generally devote modest efforts to distribute resources and assure compliance with federal and state laws. The traditional focus of state agencies—to enforce federal requirements, enact state policies, and act as a conduit for the flow of federal money to school districts—means they lack both the staff and expertise to reform schools (McDermott, 2004; Sunderman & Orfield, 2006). In such a system, responsibility for school improvement gets passed down to the next level of the educational system, often leaving low-capacity schools to improve on their own.

Coping with Impracticality

Policy recipients have two ways to counteract the impracticalities associated with a law handed down to them. They can soften the impact of the law by exploiting loopholes or weaknesses in enforcement, or they can demand design changes. State governments have moved on both fronts. The most obvious way to keep the system at bay is to maintain modest proficiency targets, to postpone higher growth expectations, or to lower proficiency cut-offs or test rigor as needed. Some have termed this tendency the “race to the bottom” (Peterson & Hess, 2008). States can tolerate or encourage the

adoption of less-invasive forms of corrective action and restructuring, another tendency that has been reported by research (Center for Education Policy, 2007, 2008c). Or states can manipulate how AYP is determined by adopting complicated statistical techniques, changing subgroups size, or setting different targets for different subgroups (Sunderman, 2006).

States have also pressured the federal government to make implementation more flexible. Indeed, the federal government has moved to attenuate some of the shortcomings of the system, but none have removed the sense of pervasive impracticality surrounding the sanctions system. This is particularly evident in two pilot programs: the growth model pilot program and the differentiated accountability pilot program. The growth model pilot was intended to allow states to take into account student progress when determining AYP, but early indications are that its use has made little difference in the number of schools identified for improvement (Klein, 2007). The differentiated accountability pilot program would allow some states to determine how to intervene in schools and districts. The rationale was straightforward: schools need “differentiated” interventions that are linked to the reasons for which they were identified for improvement in the first place. However, the pilot program continued to require states to increase the number of students participating in the supplemental services and choice options, to maintain the NCLB sanctions stages with stipulated timelines, and ensure that the restructuring and corrective action sanctions were retained (U.S. Department of Education, 2008, April 3). In short, states have shown a tendency to lessen compliance pressures while the federal government has made allowances for minor design changes that leave the system with its attendant impracticalities firmly in place.

In sum, the federal sanctions system as currently designed, with its simplistic method of determining performance indicators and setting goals and its rigid staging of sanctions, has proven to be quite impractical. In state systems with at least moderately high performance demands, NCLB has led to high numbers of failing schools that by far outstrip district and state intervention capacities. But it is not even clear if the bulk of these schools are in fact correctly classified. Most notably, the system has no practical answers for the full spectrum of student performance and learning needs, most notably for students performing far-below proficient, special needs students, and excluded marginally performing students. Moreover, it does not speak to the predicament of low-capacity schools and districts. While it may appear that the sanctions system has succeeded in fermenting a climate of reform, such ferment, in many instances, is more likely to result in unproductive turbulence than sustained school improvement.

Is the Sanctions System Valued and Legitimate?

School accountability systems are popular with politicians and the public (Dorn 1997), but run into serious acceptability problems with teachers and administrators in low-performing schools and districts, those whose response is critical for the success of school improvement. Studies conducted over the last ten years or so suggest that, while standards, measurement of performance, and consequences for low performance are

widely accepted ideas *in general*, attitudes about the high stakes nature of accountability systems are more negative (Conley & Goldman, 2000; Darling-Hammond, 2004; Firestone, Schorr and Monfils 2004; Hill and Lake 2002; Koretz, et al 1996; Leithwood, Steinberg and Jantzi, 2002; Loeb, Knapp & Elfers, 2008; Mintrop and Trujillo 2007; Spillane et al 2002). In particular, sanctions systems are met with wide skepticism at the school and district levels (Mintrop, 2004). Schools do use accountability systems as an orientation for performance, are responsive to them and their demands, and try (many strenuously) to avoid sanctions (Finnigan & Gross, 2007; Mintrop, 2004; Kelley and Protsik 1997). Some studies have found that accountability pressures have given impetus to school and district turnaround (Skrla & Scheurich, 2004; Hill and Lake 2002). Nevertheless, a positive outlook does not seem to be the modal attitude (Sunderman et al., 2004; Abrams, Pedulla, and Madaus, 2003; Louis, Febey, and Schroeder, 2005) and the few instances of school and district turnarounds do not warrant the invasive sanctions mandated by NCLB.

Accountability systems fashioned after NCLB principles violate core professional norms of educators and produce widespread frustration and de-moralization among those charged with carrying out needed school improvement efforts. While “teaching to the test” is acceptable to a certain degree, high pressure to do so, to the exclusion of other more complex and far-reaching goals, is not. As a result, teachers widely report that they need to compromise standards of good teaching when striving to meet accountability goals (Abrams, Pedulla, & Madaus, 2003; McNeil, 2000; Valenzuela 2005). Indeed, schools’ performance or accountability status may be a poor indicator of their overall educational quality (Mintrop and Trujillo, 2007).

The moral discourse of accountability assigns failure to schools’ lack of high expectations and standards for all students, and places the burden of responsibility on educators. Educators themselves are torn. They assume guilt and at the same time discount culpability (Hargreaves 2004; Mintrop 2004; Finnigan & Gross, 2007; Booher-Jennings, 2005). The belief is wide-spread that sanctions systems penalize teachers and administrators who often have to work under the most difficult conditions, in schools that serve children in poverty from many different demographic sub-groups (Sunderman et al., 2004). As a result, low-performance labels attached to the organization are widely rejected as valid judgments of individual work quality (Mintrop, 2004).

Accountability goals are often not seen as realistic, with sanctions viewed as ill-guided and of little personal consequence that unfairly place blame on teachers. Despite misgivings, the humiliation or discomfort of working in a publicly labeled low-performing school seems to trigger an initial surge of energy and determination, if not frenzy, among educators to meet the goals (Finnigan and Gross, 2007; Mintrop 2004, Fullan, 2003; Malen, Croninger, Muncey, & Redmond-Jones, 2002). Often the most activist teachers and administrators who have the least reason to own the low-performance label are the ones who assume most of the responsibility. In other cases, district and school administrators use the label to demand compliance with centrally adopted prescriptions and delegitimize teachers’ traditional defenses against administrative intrusions.

When hoped-for improvements are either not forthcoming, or cannot be sustained after the short-term fixes have been exploited (as is often the case in struggling low-capacity schools), resentment and demoralization set in and trigger exit (Finnigan and Gross, 2007; Mintrop, 2004). The reality of high rates of school failure in the more demanding states, in the face of threatened or imposed sanctions, and despite educators' efforts or willingness to comply and suspend judgment, reinforces the overall negativity of the sanctions system. Educators who are guided by the idea of public service in a difficult environment (Sennett 2006; Shamir 1990), and who do not embrace the system's self-interested performance calculus, feel especially devalued.

But values and ideas about what a good education ought to look like are not fixed, and teachers are an occupational group that is highly susceptible to external normative influences. Accountability systems, with their measured outcomes, performance targets, sanctions, and attendant programmatic prescriptions rooted in powerful ideologies of effectiveness and science may reshape values. Furthermore, the loosening connection between teachers and schools of education that traditionally inculcated novice teachers into a discourse of professionalism and progressivism, combined with the tremendous teacher turnover in urban schools, may result in teachers' growing susceptibility to the system. These changes may make sanctions-based accountability systems an overwhelming force that not only creates new work routines, but also new values which may be more aligned with the system, and, in the process, lessening educators' misgivings.

No question, we have examples of schools that have imbued accountability goals with moral purpose, function with a sense of goal integrity or good balance between external demands and internal educational values (Reeves 2000; Elmore 2004; Mintrop and Trujillo, 2008), have taken determined steps to tightly align their teaching to state assessments and have shown to be extraordinarily successful. At the present time, however, it is probably safe to say that negativism is the prevalent mood. Despite an almost twenty-year run in some states, accountability systems, and particularly NCLB, still encounter serious legitimacy and acceptability problems among the very groups that they are designed to primarily target. It is indicative in this context that, in their majority, state accountability systems prior to NCLB have either rarely used, or turned away from, high pressure and sanctions as a main lever to motivate teachers. Instead they came to emphasize mild pressure that blends oversight with capacity building over time. By contrast, under NCLB, pressure as an improvement strategy is a central feature, and schools may face severe sanctions in a rather short time, with all the concomitant problems of legitimacy.

Conclusion: Why We Retain the Sanctions System and What it Costs

The federal sanctions system is not a powerless system, but one that nevertheless is likely to fail. The system would not fall if it was simply a matter of skepticism on the part of those who need to carry out school reform, or because it was not valued by teachers or administrators in low-performing schools and districts. It would probably not fall based on the inconclusive data on the effectiveness of accountability systems since

interpretations can be slanted, and one can cling to the encouraging upswing of test scores in most state systems. But the combination of uncertain effects, loose connection to broader educational values and norms of educators, and the difficulties or impossibilities of carrying out the system day-to-day makes it a prime candidate for declaring it a failure.

But there is a way out. As long as states aim at a low-goal horizons, and more lenient options for school improvement or restructuring continue to be chosen, the architecture of NCLB will hold. But it will be a solidly Horizon 1 undertaking. The problem with such accountability systems is not that they concentrate schools on Horizon 1 challenges, but that they tend to squelch teacher activities in Horizon 2, particularly when these systems *work well*, that is, when they push educators to run a tight ship around test-driven basic skills remediation, without allowing time to deepen students' learning. And once a system has operated within the confines of Horizon 1 for a while, and educators have internalized the intellectual habits rewarded in such a system, school improvement dynamics cannot simply be switched over into Horizon 2. Thus learning gets stuck while the system succeeds. This is particularly destructive for poor students and students of color who, more so than white students, are concentrated in the schools NCLB identifies as failing.

There are two more reasons why the system may persist even when it fails. One is that, as the literature on failing organizations has shown (Meyer & Zucker, 1989), failing structures are kept in place when groups derive secondary benefits from the maintenance of those structures. Secondary beneficiaries of NCLB are those ideologically or politically committed to NCLB, those deriving economic benefit from the system (e.g., testing agencies, educational management organizations, segments of the school improvement industry), and those deriving political benefit from the system's dysfunction (e.g., politicians campaigning on a platform of educational reform).

Perhaps more important among those who are committed to educational equity is the sense that we lack credible policy alternatives to the dominant paradigm currently driving federal and state educational policy—a focus on test-based and sanctions-driven accountability. Consequently, many policy proposals focus on ways to improve NCLB. For example, accountability systems that set targets pegged to real growth, achieved by a sizable number of demographically similar high-performing Title I schools, are preferable to the current status measures. Also preferable are systems that incorporate multiple indicators of performance. While state standards are good orientations and state tests good devices for system monitoring and self-monitoring, multiple indicators of school quality are better able to cover a wider spectrum of educational goals and valued outcomes. Multiple-indicator systems have a better chance to connect to concerns for student engagement in learning and instructional quality. There are a number of good suggestions that if adopted could improve on the meaningfulness of accountability systems for educational practice but fail to address the underlying flaws of NCLB. We argue that what is needed is an alternative to the current sanction system and a broadening of the social welfare agenda.

Recognizing the complexity of schooling, the need for capacity building, and the importance of factors external to the school that affect student performance is one alternative to the current sanctions-based accountability system. Policymakers would be required to think about school reform differently and acknowledge that schools alone cannot overcome the social and economic inequities in our society that contribute to unequal educational outcomes. More comprehensive investments in student welfare that link education with health, job development, and community building, as well as redistributive investments to attract and keep top-flight professionals in schools for the poor would be paramount.

- Overreliance on sanctions can be reduced when policies aim to develop a partnership between government, teachers, and parents, and motivate changes by adhering to the professional values and standards of educators.

But even more comprehensive social welfare measures do not by themselves solve the problems of the adversarial sanctions system currently in place. Alternatives to sanctions-driven accountability are policies based on a partnership between government, the teaching profession, and parents that motivate changes by adhering to the professional values and standards of educators. Schools cannot be improved against the better judgment, and without the enthusiastic participation, of those charged with making the improvements. While this commitment cannot be coerced through sanctions, it can be motivated through guidance and mild and positive pressure that mobilize internal ideals and standards of competence and care. For educators, such standards need to be developed through professional socialization in teacher-preparation programs and sustained by way of good instructional supervision, learning communities at school sites, professional networks – and the soft power of accountability systems that are redesigned to inspire educators. Accountability systems inspire educators when they connect to broader educational values and give the stronger teachers enough flexibility to model best practices. Soft accountability is powerfully augmented when parents are mobilized to support their children’s achievement and press for high-quality schools. We submit that after about fifteen years of state and federal sanctions-driven accountability that has yielded relatively little, it is time to try a new approach. The hard work of broader-based movements, nourished by government and civic action, will have to replace legal-administrative enforcement and mandates as the centerpiece of such an equity agenda.

REFERENCES

- Abedi, J. (2004). The No Child Left Behind Act and English language learners: Assessment and accountability issues. *Educational Researcher*, 33(1), 4-14.
- Abrams, L., Pedulla, J., & Madaus, G. (2003). Views from the classroom: teachers' opinions of statewide testing programs. *Theory into Practice*, 42 (1).
- Amrein, A. L., & Berliner, D. C. (2002). High-stakes testing, uncertainty, and student learning. Retrieved August 22, 2003, from Education Policy Analysis Archives, 10 (18) <http://epaa.asu.edu/epaa/v10n18/>
- Anagnostopoulos, D. (2003). The new accountability, student failure, and teachers' work in urban high schools. *Educational Policy*, 17(3), 291-316.
- Argyris, C. (1990). *Overcoming Organizational Defenses*. New York: Allyn and Bacon.
- Asimov, N. (2008, September 5). State falling way behind No Child Left Behind. *San Francisco Chronicle*, pp. A-1.
- Au, W. (2007). High-stakes testing and curricular control: A qualitative metasynthesis. *Educational Researcher*, 36 (5), 258-267.
- Aucoin, P., & Heintzman, R. (2000). The dialectics of accountability for performance in public management reform. *International Review of Administrative Sciences*, 66(1), 45-55.
- Balfanz, R., & Legters, N. E. (2004). Locating the dropout crisis: Which high schools produce the nation's dropouts? In G. Orfield (Ed.), *Dropouts in America: Confronting the graduation rate crisis*. (pp. 57-84). Cambridge, MA: Harvard Education Press.
- Balfanz, R., & Legters, N. E. (2008). NCLB and reforming the nation's lowest-performing high schools. In G. L. Sunderman (Ed.), *Holding NCLB accountable: Achieving accountability, equity, & school reform*. (pp. 191-207). Thousand Oaks, CA: Corwin Press.
- Booher-Jennings, J. (2005). Below the bubble: "educational triage" and the Texas accountability system. *American Educational Research Journal*, 42 (2), 231-268.
- Batt, L., Kim, J., & Sunderman, G. L. (2005). *Limited English proficiency students: Increased accountability under NCLB*. Cambridge, MA: The Civil Rights Project at Harvard University.
- Buchanan, B. (2006). *Turnover at the Top: Superintendent Vacancies and the Urban School*. Lanham, PA: Rowman & Littlefield.
- Carnoy, M., & Loeb, S. (2002). Does external accountability affect student outcomes? *Education Evaluation and Policy Analysis*, 24, 305-332.
- Center for Education Policy. (2007). *Moving beyond identification: Assisting schools in improvement*. Washington, DC: Author.
- Center for Education Policy. (2008a). *A call to restructure restructuring: Lessons from the No Child Left Behind Act in five states*. Washington, DC: Author.
- Center for Education Policy. (2008b). *Has student achievement increased since 2002? State test score trends through 2006-07*. Washington, DC: Author.
- Center for Education Policy. (2008c). *Managing more than a thousand remodeling projects: School restructuring in California*. Washington, DC: Author.
- Chicago Public Schools. (2007). *SES tutoring programs: An evaluation of year 3 in the Chicago Public Schools*. Chicago: Author.

- Coleman, J.S. (1990). *Foundations of Social Theory*. Cambridge: Harvard University Press, Belknap Press.
- Coltrane, B. (2002, November). English language learners and high-stakes tests: An overview of the issues. Retrieved May 13, 2004, from <http://www.cal.org/ericell/DIGEST>
- Conley, D., & Goldman, P. (2000). Half full, half empty? Educator responses over time to state-level systematic reforms. *International Journal of Educational Reform*, 9 (3) 249-60.
- Cooper, C.L., & Robertson, I. (Eds.), *International review of industrial and organizational psychology*. Chichester, England: Wiley, 1986
- Darling-Hammond, L. (2004). Standards, accountability, and school reform, *Teachers College Record*, 106 (6), 1047-85.
- Diamond, J. B., & Spillane, J. (2004). High-stakes accountability in urban elementary schools: Challenging or reproducing inequality? *Teachers College Record*, 106(6), 1145-1176.
- Education Commission of the States. (2004, March). State takeovers and reconstitutions. Retrieved March 28, 2006, from <http://ecs.org/clearinghouse/51/67/5167.htm>
- Elmore, R. F. (2004). *School reform from the inside out: Policy, practice and performance*. Cambridge, MA: Harvard Education Press.
- Finnigan, K., & Gross, B. (2007). Do accountability policy sanctions influence teacher motivation? Lessons from Chicago's low-performing school. *American Educational Research Journal*, 44 (3), 594-629.
- Firestone, W., Monfils, L., & Schorr, R. (2004). *The Ambiguity of Teaching to the Test: Standards, Assessment, and Educational Reform*. New Jersey: Lawrence Erlbaum Associates, Inc., Publishers.
- Frey, B.S. (1997). On the relationship between intrinsic and extrinsic work motivation. *International Journal of Industrial Organization*, 15 (4), 427-440.
- Fullan, M. (2003). *Change forces with a vengeance*. London, New York: Routledge Farmer.
- Gill, B., Zimmer, R., Christman, J., & Blanc, S. (2007). *State takeover, school restructuring, private management, and student achievement in Philadelphia*. Santa Monica: Rand.
- Haney, W. M. (2000). The myth of the Texas miracle in education. Retrieved August 22, 2006, from Education Policy Analysis Archives, 8 (41) <http://epaa.asu.edu/epaa/v8n41/>
- Haney, W. M. (2008). Evidence on education under NCLB (and how Florida boosted NAEP scores and reduced the race gap). In G. L. Sunderman (Ed.), *Holding NCLB accountable: Achieving accountability, equity, & school reform*. (pp. 91-102). Thousand Oaks, CA: Corwin Press.
- Hanushek, E. A., & Raymond, M. E. (2004). Does school accountability lead to improved performance? *Journal of Policy Analysis and Management*, 24(2), 297-327.
- Hargreaves, A. (2004). Distinction and disgust: The emotional politics of school failure. *International Journal of Leadership in Education*, 7 (1), 27-41.

- Heinrich, C. J., Meyer, R. H., & Whitten, G. (2008, March). *Supplemental educational services under No Child Left Behind: Who signs up, and what do they gain?* Paper presented at the American Education Research Association, New York, NY.
- Heistad, D. (2006). *Analysis of 2005 supplemental educational services in Minneapolis Public Schools: An application of matched sample statistical design*. Minneapolis: Minneapolis Public Schools, Research, Evaluation and Assessment Department.
- Herman, J. L. (2004). The effects of testing on instruction. In S. Fuhrman & R. Elmore (Eds.), *Redesigning accountability* (pp. 141–146). New York: Teachers College Press.
- Hill, P., Lake, R., & Celio, M. (2002). *Charter schools and accountability in public education*. Washington, DC: Brookings Institution Press.
- Hoff, D. J. (2007). Provision on tutoring raises renewal issues. *Education Week*, 27, 1, 19.
- Johnson, S., & Birkeland, S. (2003). Pursuing a “sense of success”: New teachers explain their career decisions. *American Educational Research Journal*, 40 (3), 581-617.
- Kelley, C., & Protsik, J. (1997). Risk and reward: Perspectives on the implementation of Kentucky’s school-based performance award program. *Educational Administration Quarterly*, 33 (4), 474-505.
- Kieffer, M. J., Lesaux, N. K., & Snow, C. E. (2008). Promises and pitfalls: Implications of NCLB for identifying, assessing, and educating English language learners In G. L. Sunderman (Ed.), *Holding NCLB accountable: Achieving accountability, equity, & school reform* (pp. 57-74). Thousand Oaks, CA: Corwin Press.
- Kim, J., & Sunderman, G. L. (2004). *Does NCLB provide good choices for students in low performing schools?* Cambridge: The Civil Rights Project at Harvard.
- Kim, J., & Sunderman, G. L. (2005). Measuring academic performance under NCLB: Implications for educational equity. *Educational Researcher*, 34(8), 3-13.
- Klein, A. (2007, December 19). Impact is slight for early states using 'growth'. *Education Week*, pp. 24-25,
- Koretz, D. M. (2008). The pending reauthorization of NCLB: An opportunity to rethink the basic strategy. In G. L. Sunderman (Ed.), *Holding NCLB accountable: Achieving accountability, equity, & school reform*. (pp. 9-26). Thousand Oaks, CA: Corwin Press.
- Latham, G., & Pinder, C. (2005). Work motivation theory and research at the dawn of the twenty-first century. *Annual Review of Psychology*, 56 (1), 485-516.
- Lawler, E. (1973). *Motivation in work organizations*. San Francisco: Jossey-Bass.
- Lee, J. (2007). *The testing gap: Scientific trials of test-driven school accountability systems for excellence and equity*. Charlotte, NC: Information Age Publishing.
- Lee, J., & Wong, K. K. (2004). The impact of accountability on racial and socioeconomic equity: Considering both school resources and achievement outcomes. *American Educational Research Journal*, 41(4).
- Leithwood, K., Steinbach, R., & Jantzi, D. (2002). School Leadership and Teachers' Motivation To Implement Accountability Policies. *Educational Administration Quarterly*, 38 (1) 94-119.
- Linn, R. L. (2005). Conflicting demands of No Child Left Behind and state systems: Mixed messages about school performance. *Education Policy Analysis Archives*, 13(33).

- Linn, R. L. (2008). Toward a more effective definition of adequate yearly progress. In G. L. Sunderman (Ed.), *Holding NCLB accountable: Achieving accountability, equity, & school reform* (pp. 27-42). Thousand Oaks, CA: Corwin Press.
- Loeb, H., Knapp, M. & Elfers, A. (2008). Teachers' response to standards-based reform: Probing reform assumptions in Washington State. *Education Policy Analysis Archives*, 16 (8), 1-29.
- Lortie, D. C. (1977). *Schoolteacher: A sociological study* (Phoenix ed.). Chicago, IL: University of Chicago Press.
- Louis, K.S., Febey, K., & Schroeder, R. (2005). State-mandated accountability in high schools: Teachers' interpretations of a new era. *Educational Evaluation and Policy Analysis*, 27, 177–204.
- Malen, B., Croninger, R., Muncey, D., & Redmond-Jones, D. (2002). Reconstituting schools: "Testing" the "theory of action." *Educational Evaluation and Policy Analysis*, 24 (2), 113-132.
- Marion, S. T., White, C., Carlson, D., Erpenbach, W. J., Rabinowitz, A., & Sheinker, J. (2002). *Making valid and reliable decisions in determining adequate yearly progress*. Washington, DC: Council of Chief State School Officers.
- McDermott, K. A. (2004). Incentives, capacity, and implementation: Evidence from Massachusetts education reform. *Journal of Administration Research and Theory*, 16, 45-65.
- McNeil, L. (2000). *Contradictions of school reform*. New York, London: Routledge.
- Meyer, M., & Zucker, L. (1989). *Permanently failing organizations*. Thousand Oaks, CA: Sage.
- Mintrop, H. (2004). *Schools on probation: How accountability works (and doesn't work)*. New York: Teachers College.
- Mintrop, H., & Trujillo, T. (2005). Corrective action in low performing schools: Lessons for NCLB implementation from first-generation accountability systems. *Education Policy Analysis Archives*, 13(48).
- Mintrop, H. & Trujillo, T. (2007). The practical relevance of accountability systems for school improvement: A descriptive analysis of California schools. *Educational Evaluation and Policy Analysis*, 29 (4), 319-352.
- Mintrop, H. & Trujillo, T. (2008). High-stakes accountability and teacher learning community -- How do they go together. Paper presented at the Annual Conference of the University Council for Educational Administration (UCEA), Orlando, FL.
- Moore-Johnson, S. (2007). *Finders and keepers: Helping new teachers survive and thrive in our schools*. Indianapolis, IN: Jossey-Bass.
- Neal, D., & Schanzenbach, D. W. (2008). Left behind by design: Proficiency counts and test-based accountability. University of Chicago.
- Novak, J. F., B. (2003, December). *Penalizing diverse schools? Similar test scores but different students bring federal sanctions*. Berkeley, CA: Policy Analysis for California Education.
- Owens, A., & Sunderman, G. L. (2006). *School accountability under NCLB: Aid or obstacle for measuring racial equity?* Cambridge, MA: The Civil Rights Project at Harvard University.

- Peterson, P. E., & Hess, F. M. (2008). Few states set world-class standards: In fact, most render the notion of proficiency meaningless. *Education Next*, 8(3), 70-73.
- Posner, R.A., & Rasmusen, E. B. (1999). Creating and enforcing norms, with special reference to sanctions. *International Review of Law and Economics*, 19 (3), 369-382.
- Potter, A., Ross, S. M., Paek, J., McKay, D., Ashton, J., & Sanders, W. L. (2007). *Supplemental educational services in the state of Tennessee: 2005-06 (2004-2005 student achievement results)*. Memphis, TN: Center for Research in Education Policy, University of Memphis.
- Raymond, M. E., & Hanushek, E. A. (2003). High-stakes research: Accountability works after all. *Education Next*, 3, 48-55.
- Reeves, D. (2000). *Accountability in action*. Denver: Advanced Learning Press.
- Rickles, J. H., Barnhart, M. K., & Gualpa, A. S. (2008). *Supplemental educational services participation and impact on student achievement: The case of one urban district over five years*. Paper presented at the American Education Research Association.
- Saltman, K.J. (2005). *The Edison Schools*. New York: Routledge
- Shamir, B. (1990). Calculations, values, and identities: The sources of collectivistic work motivation. *Human Relations*, 43 (4), 313-332.
- Shamir, B. (1991). Meaning, Self and Motivation in Organizations. *Organization Studies*, 12 (3), 405-424
- Sennett, R. (2006). *Culture of New Capitalism*. New Haven: Yale University Press
- Skrla, L., & Scheurich, J. (2004). Displacing deficit thinking in school district leadership. In Skrla, L., & Scheurich, J. (Eds.), *Educational Equity and Accountability*. (pp.109-132). New York: Routledge Farmer.
- Staw, B.M., Lance, E., & Dutton, J. (1981). Threat-rigidity effects in organizational behavior: A multilevel analysis. *Administrative Science Quarterly*, 26 501-524.
- Sunderman, G. L. (2001). Accountability mandates and the implementation of Title I schoolwide programs: A comparison of three urban districts. *Educational Administration Quarterly*, 37(4), 503-532.
- Sunderman, G. L. (2006). *The unraveling of No Child Left Behind: How negotiated changes transform the law*. Cambridge, MA: The Civil Rights Project at Harvard University.
- Sunderman, G. L., Kim, J. S., & Orfield, G. (2005). *NCLB meets school realities: Lessons from the field*. Thousand Oaks, CA: Corwin Press.
- Sunderman, G. L., & Orfield, G. (2006). Domesticating a revolution: No Child Left Behind and state administrative response. *Harvard Educational Review*, 76(4), 526-556.
- Sunderman, G. L., Tracey, C. A., Kim, J., & Orfield, G. (2004). *Listening to teachers: Classroom realities and No Child Left Behind*. Cambridge, MA: The Civil Rights Project at Harvard University.
- The Education Trust. (2004). *Measured progress: Achievement rises and gaps narrow, but slowly*. Washington, DC: Author.
- The Education Trust. (2005). *Stalled in secondary: A look at student achievement since the No Child Left Behind Act*. Washington, DC: Author.

- U.S. Department of Education. (2008, April 3). Peer review guidance for the NCLB differentiated accountability pilot applications. Retrieved July 2, 2008, from <http://www.ed.gov/policy/elsec/guid/daguidance.doc>
- Valenzuela, A. (2005). Leaving children behind: How "Texas-style" accountability fails Latino youth. Albany: State University of New York Press.
- Zimmer, R., Gill, B., Razquin, P., Booker, K., & Lockwood, J. R. (2007). *State and local implementation of the No Child Left Behind Act: Volume I--Title I school choice, supplemental educational services, and student achievement*. Washington, DC: RAND.