



# Predictable Failure of Federal Sanctions-Driven Accountability for School Improvement—And Why We May Retain It Anyway

Heinrich Mintrop and Gail L. Sunderman

The federal accountability system, made universal through the No Child Left Behind Act of 2002, is a system driven by quotas and sanctions, stipulating the progression of underperforming schools through sanctions based on meeting performance quotas for specific demographic groups. The authors examine whether the current federal accountability system is likely to succeed or fail, by asking, Does the sanctions-driven accountability system work? Is it practical? And is it legitimate among those who must implement it? The authors argue that even though sanctions-driven accountability may fail on practical outcomes, it may be retained for its secondary benefits and because there is a sense that credible policy alternatives are lacking. They conclude by proposing alternative policies and approaches to the current system.

**Keywords:** accountability; educational policy; educational reform; policy

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 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. Time and again, 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). The “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 and with up-to-date skills can reap enormous rewards, whereas 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. Although 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, although 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 1990s and then made universal through the federal No Child Left Behind (NCLB) legislation. 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 for the middle class, they 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 (Sennett, 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 management system increases central control by top management, freezing out the mediating functions of the middle layers of the organization. In the political realm, this new approach to public administration increases the potential for 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 central management of school performance 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. Whereas in private industry the new performance management system 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.

Current accountability systems in education consist of standards as broadly framed orientations for subject matter content and skills, standardized tests as the basis for performance indicators, and performance 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 the central performance demands seriously.

At its heart, the federal accountability system for schools is driven by quotas and sanctions. The substance of academic content, testing rigor, the regulation of inputs (with the exception of the requirement for highly qualified teachers), the specific method of school or program improvement, the specific methods of restructuring, all this is left for states to decide, as long as the elements 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: from identification and publication of “school improvement” status (a kind of public shaming with potentially far-reaching market consequences), to mounting loss of organizational autonomy through required external intervention and service contracting, and finally to termination through reorganization or takeover of the organization. Even though the law formulates the sanction stages 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 to 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 an approach, 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 contours of the failure of the federal sanctions-driven approach have come into appearance, as the simplicity of the remedy has become increasingly submerged by the complexity of school and district improvement. But because it is seen as an “easy” way to make the system work, it may be retained anyway.

### **The Power of Sanctions**

“Accountability is here to stay,” a slogan often heard in the early 1990s to persuade disbelieving teachers to make the necessary adjustments, was not so wrong after all. State accountability

systems have shown staying power—certainly beyond what critics imagined. And by all indications, they have proved to be quite powerful instruments in reshaping how schools go about their business, especially schools for poor and minority children (Au, 2007; Herman, 2004). Accountability systems divide into two main components: guidance in the form of standards and test data that orient instruction and inform performance; 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 being more informative (Frey, 1997). Sanctions are penalties for noncompliance 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 can be quite costly, as the outlays for the criminal justice system in the United States demonstrate. Costs are reduced when the actual imposition of sanctions is the exception because 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,” that is, 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 are likely to fail when they produce ambiguous or uncertain outcomes so that their effect is in doubt. They are likely to fail when they are designed with less-than-credible threats, aimed at diffuse target actors, and beset with visible enforcement challenges, and when they require 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, Lance, & Dutton, 1981) that prevents organizational actors from learning and trying out new solutions. Nevertheless, the approach may still be maintained (Meyer & Zucker, 1989) as long as secondary benefits can be derived from it, such as symbolic or ideological satisfaction or economic gain or simply because of the lack of realistic alternatives that address the problem as forcefully as the sanctions approach, on the face of it, promises to do.

We argue in this article that we have reached a fork in the road: The federal sanctions-driven approach to school performance has developed the adverse conditions enumerated above, in which case its failure becomes predictable. It may be fixed or made to work without real success, in which case it survives but its survival becomes educationally undesirable; or it may hang 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: (a) Does the sanctions-driven

approach work; that is, does it produce the intended results? (b) Is it practical; that is, can it be implemented? and (c) Is it valued and legitimate?

### *Does the Sanctions-Driven Accountability 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, the disruption created by restructuring, the loss of control of local actors over their affairs, and the neglect of nonmeasurable educational goals and values. This is doubtful.

*Does the system produce the expected outcomes?* To date, scholars argue whether the law has had the powerful impact on schools that it was designed for. 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, 2006, 2007, 2008; 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, 2008). 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 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; The Education Trust, 2004, 2005; Neal & Schanzenbach, 2007). But the picture appears 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 looks better for elementary grades than secondary ones, and better for math than for reading—reading scores tend to remain flat, whereas Grade 4 math scores showed some improvement. A meta-analysis of large-scale studies using national data showed modestly positive policy effects on average but no significant effect on narrowing the achievement gap (Lee, 2008). Nonetheless, there are substantial variations among states, and few states have narrowed the achievement gap among racial and socioeconomic subgroups and improved overall performance at the same time (Lee, 2007).

Although 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 the NAEP and state assessment results, it is not quite clear what the state tests measure. By all indications, 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 those in 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 that 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 the transfer option produces improvements in school performance is a moot point because the percentage of students taking advantage of the option (about 1% of eligible students) is too low to reliably attribute school improvement effects to it. More students have taken advantage of supplemental educational services (14% of eligible students in 2005–2006; 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 in 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 takeover by education management organizations (EMOs), may work in some situations but do not appear to work across the board and are often accompanied by negative side effects (see Mintrop & Trujillo, 2005, for more detail; Mathis, 2009). For example, in Maryland, 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, Croninger, Muncey,

& Redmond-Jones, 2002). Results from reconstitutions in Chicago (Hess, 2003) and in New York's Schools Under Registration Review program were inconclusive as well (Brady, 2003; New York State Education Department, 2003). School takeover by EMOs has worked in some cases but not in others, as research from Baltimore and Philadelphia suggests (Blanc, 2003; Bracey, 2002; Saltman, 2005; Travers, 2003a; Useem, 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 takeover by states (Garland, 2003). However, equally dramatic *academic* success has been much harder to achieve (Education Commission of the States, 2004; Ziebarth, 2002). Although 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 (The Brown Center, 2003; Gill, Zimmer, Christman, & Blanc, 2007). Charter schools also tend to show up on states' lists of failing schools in larger proportions than regular public schools (The Brown Center, 2003).

Thus among the variety of corrective action and restructuring strategies that have been tried, none stand out as universally effective or robust enough to overcome the power of local context. Competence of provider personnel, intervention designs, political power of actors in the system, and district and site organizational capacity to absorb the strategies all strongly influence how a particular strategy will turn out. A recent review of the available research evidence on the NCLB restructuring options study corroborates this conclusion (Mathis, 2009).

The reluctance of states and districts to embrace the strong tools provided to them by NCLB (Center for Education Policy, 2007, 2008c; U.S. Government Accountability Office, 2007) 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 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 law. Finally, if the NCLB sanctions "worked," 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-driven approach would end even if a causal connection between specific sanctions and test scores could not be found. But there is not clear-cut evidence. The 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 the effect of NCLB on student achievement—uncertainties that undermine a justification for the costs associated with the law. For those

committed to the idea of sanctions-based accountability, 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 wisdom of sanctions-based systems, but it does not silence those with second thoughts.

### *Is the Sanctions-Driven System Practical?*

In asking this question, we are less concerned with effects and more with processes of productive school improvement that NCLB is designed to leverage. If the system were practical, it would identify with high accuracy schools in need of improvement and restructuring; appropriately direct schools to pay attention to students most in need of help; produce an intervention burden for states and districts commensurate with their capacities to provide new impetus, ideas, resources, and personnel; and, 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 by those practicality criteria.

*Simplistic goal setting and misidentification of schools.* Evidence has accumulated that federal goal setting in terms of adequate yearly progress (AYP) toward proficiency has created much movement around the cutoff points but is unable to register growth in the below-proficiency band where many high-poverty schools operate (Neal & Schanzenbach, 2007); is insensitive to the exclusion of students (Amrein & Berliner, 2002; Balfanz & Legters, 2008; Haney, 2008); creates an undue burden for schools with highly heterogeneous student populations, particularly schools with large numbers of special needs students and many subgroups (Sunderman, Kim, & Orfield, 2005); and results in an untenable number of failing and misidentified schools in all but the most lenient state systems.

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—as currently defined and used in most states—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 that are identified for improvement and schools that met the federal AYP goals. For example, an analysis of data from Virginia found similar levels of improvement in proficiency in both types of schools (Kim & Sunderman, 2005). (Whether or how the growth model pilot program now allowed under NCLB may attenuate this problem is discussed below.) If AYP, as currently used in most states, were one among many school quality indicators used to gauge school and district performance, these

inaccuracies would be tolerable, perhaps, but it is not. As the sole authoritative indicator that triggers sanctions, it creates powerful realities for schools and districts on the ground. Irrationally, many schools may be forced to own the low-performance label when they are in fact fairly healthy and still making progress.

*Insensitivity to exclusion.* Given that extraschool factors have a large effect on variation in student achievement and school performance (Scheerens & 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 are not performing at the proficient level of academic achievement and do not graduate 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). Exclusion—for example retention of students prior to key testing grades, reassigning weak learners to nonaccountable categories, or pushing students out of school (Anagnostopoulos, 2003; Booher-Jennings, 2005; Haney, 2008)—is one of various coping strategies 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 straightforward 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, there are at least 5 targets and there could be as many as 37 (Marion et al., 2002), increasing the likelihood of failure in proportion to the number of subgroups. 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 subgroup targets for students with limited English proficiency (LEP) or 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: 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 have questioned the fairness of the provisions because students who achieve 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 have found that schools reporting an LEP subgroup are more likely to be identified as needing improvement than those without this subgroup—a pattern that applies to the subgroup of students with disabilities as well (Batt et al., 2005; Sunderman et al., 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 short-changed, as it is in schools' and districts' self-interest to ration their services according to the most pressing accountability demands.

*Discouraging rigor.* Fundamentally, the practicality of sanctions-driven accountability is a function of the testing rigor of state accountability systems. Michael Fullan's (2003) idea of goal horizons, 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 prescriptions, improvement of basic organizational effectiveness, and all sorts of short-term strategies—such as targeting students at statistical cutoff points, teaching to the test, standard-specific remediation, pep rallies at testing time, and so on—can have an initial significant effect on school improvement as measured by test scores. But the gains do not seem to persist (Fullan, 2003). In 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. Although 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 suffice to reach these higher order goals.

State accountability systems that operate within Horizon 1 seem to be more practical in 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 have capacity-building needs that are fairly light, affordable, and manageable (Peterson & Hess, 2006). Pressure may be sufficient to prod schools toward reaching the system's modest proficiency goals. Thus in low-demand or low-rigor systems, a mere sanctions approach of the NCLB type would work. In contrast, we have systems that are more ambitious in their performance demands while at the same time producing an intervention burden that seems to make the system *unworkable*. To illustrate, data from the 2003–2004 school year (Mintrop, 2008; Mintrop & Trujillo, 2005) show that states such as California, with high testing rigor in which the definition of proficiency is relatively close to that of NAEP, 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—about 5% of the schools in Texas were identified as needing improvement. Kentucky is a state with medium testing rigor and a correspondingly medium intervention burden (about 10% of total number of schools in 2003–2004).

States that adopted accountability systems prior to NCLB (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% 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 possible under NCLB, with its 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 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, Useem, Travers, & Lesnick, 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 (Moore-Johnson, 2007) and do not 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 and more if they will get them wet in the first place. None of the reorganization measures commonly associated with federal (and state) sanctions address 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), and has already happened in districts that have 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 do regular public schools. For these schools as well, fundamental redesign happened prior to the introduction of federal sanctions.

In other words, rather than being distinct stages of well-articulated intervention intensity, 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 federal sanctions makes designing measures appropriate to the developmental needs of a given school or district much more difficult.

*Strained capacities.* The impracticality of NCLB 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 situation of course is more likely 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 and 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.

In contrast, many poorly performing schools lack the resources and capacity to respond, on their own, to sanctions in ways that will improve curriculum and instruction (Elmore, 2004). Intensifying pressure through sanctions will not result in improvements but in further rigidity, fragmentation, and deterioration (Mintrop, 2004). Low-capacity schools are predestined to bank on short-term strategies that require little added capacity (Sunderman, Tracey, Kim, & Orfield, 2004). Common strategies are 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 the mere threat of sanctions is not strong enough to lead 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 include disaggregation by subgroup categories and teacher quality, which is more extensive than previous data requirements. Even more important, states have a role in helping schools and districts improve under NCLB, a requirement that traditionally has not been a state function. State education agencies are relatively small and 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 conduits for the flow of federal money to school districts—means they lack both the staff and the 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 sanctions at bay is to maintain modest proficiency targets, to postpone higher growth expectations, or to lower proficiency cutoffs 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 put pressure on the federal government to make implementation more flexible. And indeed, the federal government has moved to attenuate some of the shortcomings of NCLB but has not removed the sense of pervasive impracticality surrounding the sanctions approach. This is particularly evident in two pilot programs—the growth model pilot program and the differentiated accountability pilot program. As of January 2009, the U.S. Department of Education had approved the growth model pilot program in 15 states. This pilot program was intended to allow states to take into account student progress when determining AYP, but early, although not definite, indications are that its use has made little difference in the number of schools identified for improvement (Klein, 2007; Weiss, 2008a, 2008b). Under current guidelines the pilot maintains many of the requirements of the current status measure—the 100% proficiency requirement remains in effect, states are required to incorporate grade-level proficiency targets into the model, and only assessments in reading and math are included—and will do little to mitigate misidentification and the enormous intervention burden that accumulates in states with medium to high standards. Moreover, growth models are challenging to design, requiring very sophisticated state

databases and a high level of technical expertise (Haertel, 2005; McCaffrey, Lockwood, Koretz, & Hamilton, 2003), both in short supply in the majority of states.

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 time lines, and to ensure that the restructuring and corrective action sanctions were retained (U.S. Department of Education, 2008). In short, states have shown a tendency to lessen compliance pressures, whereas the federal government has made allowances for minor design changes that leave the overall approach with its attendant impracticalities firmly in place.

In sum, the federal sanctions-driven approach to school performance as currently designed with its simplistic method of determining performance indicators and setting goals and its rigid staging of sanctions has proved 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 sanctions system has no practical answers in stock for the full spectrum of student performance and learning needs, especially for students 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. Although it may appear that the pressures of looming sanctions have succeeded in fermenting a climate of reform, such ferment, in many instances, is more likely to result in unproductive turbulence than in sustained school improvement.

### *Is the Sanctions-Driven System Valued and Legitimate?*

School accountability systems are popular with politicians and the public (Dorn, 1998) but run into serious acceptability problems with those whose response is critical for the success of school improvement—the teachers and administrators in low-performing schools and districts. Studies conducted over the past 10 years or so suggest that, although 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, & Monfils, 2004; Hill, Lake, & Celio, 2002; Koretz, Barron, Mitchell, & Stecher, 1996; Leithwood, Steinberg, & Jantzi, 2002; Loeb, Knapp, & Elfers, 2008; Mintrop & Trujillo, 2007; Spillane et al., 2002). In particular, sanctions 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 (Kelley & Protsik, 1997; Mintrop, 2004). Some studies have found that accountability pressures have given impetus to school and district turnaround (Hill et al., 2002; Skrla & Scheurich, 2004). Nevertheless, a positive outlook does not seem to be the modal attitude (Abrams, Pedulla, & Madaus,

2003; Louis, Febey, & Schroeder, 2005; Sunderman et al., 2004), 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 demoralization among those charged with carrying out needed school improvement efforts. Although 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 et al., 2003; McNeil, 2000; Valenzuela, 2005). Indeed, schools' performance or accountability status may be a poor indicator of their overall educational quality (Mintrop & 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 it (Booher-Jennings, 2005; Finnigan & Gross, 2007; Hargreaves, 2004; Mintrop, 2004). The belief is widespread that sanctions penalize teachers and administrators who have to work under the most difficult conditions in schools that serve children in poverty from many different demographic subgroups, a belief that resonates with evidence documented by research (Sunderman et al., 2004). As a result, low-performance labels attached to the organization are rejected as valid judgments of individual work quality (Mintrop, 2004).

Studies have documented that accountability goals are not seen as realistic, and sanctions are seen as ill guided and of little personal consequence, unfairly placing blame on teachers. Yet 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 & Gross, 2007; Fullan, 2003; Malen et al., 2002; Mintrop, 2004). 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 either are 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 that trigger exit (Finnigan & 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-driven approach. 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 are not fixed, and teachers as an occupational group are 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. We have examples of schools that have imbued accountability goals with moral purpose and that function with a sense of goal integrity, or good balance between external demands and internal values (Elmore, 2004; Mintrop & Trujillo, 2008; Reeves, 2000); have taken determined steps to tightly align their teaching to state assessments; and have been extraordinarily successful. 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 dwindling teacher agency and make accountability systems and sanctions-driven accountability an overwhelming force that creates not only new work routines but also new values.

At the present time, however, it is probably safe to say that negativism is the prevalent mood. And despite an almost 20-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 established 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. 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

The federal sanctions-driven approach to school performance is not a powerless system but is nevertheless likely to fail. The system would not fail if it were 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 fail based on the inconclusive data on the effectiveness of accountability systems, because 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 the broader educational values and norms of educators, and the difficulties or impossibilities of carrying out the law's regulations day-to-day makes it a prime candidate for being declared a failing system.

But there is a way out. As long as states maintain a low goal horizon and the 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. 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 that 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 it, those deriving economic benefit from it (e.g., testing agencies, educational management organizations, segments of the school improvement industry), and those deriving political benefit from the dysfunction of the sanctions-driven approach (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. A focus on sanctions-driven accountability is the dominant paradigm that is currently driving federal and state educational policy, and consequently many policy proposals focus on ways to improve on NCLB. Some of these proposals could 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. Although state standards are good orientations and state tests are good devices for system monitoring and self-monitoring, multiple indicators of school quality are better able to cover a wide spectrum of educational goals and valued outcomes. Multiple-indicator systems have a better chance to connect to concerns such as 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 sanctions approach 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. Policy makers 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.

But even more comprehensive social welfare measures do not by themselves solve the problems of the adversarial sanctions system currently in place. An alternative approach to sanctions-driven accountability would be a partnership between government, the teaching profession, and empowered low-income parents. Schools cannot be improved against the better judgment, and without the enthusiastic participation, of those charged with making the improvements. Their commitment cannot be coerced through sanctions, but it can be motivated through guidance and mild and positive pressures that mobilize internal ideals and standards of competence and care (Darling-Hammond, 2004). Such standards, pertaining to teachers' inherent desire to reach students personally and effect learning, are exemplified by the

actions of the stronger teachers and need to be made compelling for all teachers 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. In such systems, state tests become a feedback device rather than an automatic sanctions trigger. 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 15 years of state and federal sanctions-driven accountability that has yielded relatively little, it is time to try a new approach, one that centers on the idea of sharing responsibility among government, the teaching profession, and low-income parents. The hard cultural 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.

#### NOTE

This research was supported in part by a grant from the Charles Stewart Mott Foundation to the Civil Rights Project/Proyecto Derechos Civiles, University of California, Los Angeles.

#### 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), 18–29.
- Amrein, A. L., & Berliner, D. C. (2002). High-stakes testing, uncertainty, and student learning. *Education Policy Analysis Archives*, 10(18), Retrieved August 22, 2003, from <http://epaa.asu.edu/epaa/v10n18/>
- Anagnostopoulos, D. (2003). The new accountability, student failure, and teachers' work in urban high schools. *Educational Policy*, 17, 291–316.
- Argyris, C. (1990). *Overcoming organizational defenses*. New York: Allyn & Bacon.
- Asimov, N. (2008, September 5). State falling way behind No Child Left Behind. *San Francisco Chronicle*, p. A-1.
- Au, W. (2007). High-stakes testing and curricular control: A qualitative metasynthesis. *Educational Researcher*, 36, 258–267.
- 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, and school reform* (pp. 191–207). Thousand Oaks, CA: Corwin Press.
- 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.
- Blanc, S. (2003, Summer). Principals offer mixed perspectives on first year of takeover: Supports and challenges in the multiple provider model. *The Notebook*. Retrieved June 7, 2009, from <http://www.thenotebook.org/editions/2003/summer/principals.htm>

- Booher-Jennings, J. (2005). Below the bubble: "Educational triage" and the Texas accountability system. *American Educational Research Journal*, 42, 231–268.
- Bracey, G. (2002). *The market in theory meets the market in practice: The case of Edison schools*. Tempe, AZ: The Education Policy Studies Laboratory at Arizona State University. Retrieved November 2, 2005, from <http://www.epicpolicy.org/files/EPsL-0202-107-EPRU.pdf>
- Brady, R. (2003). *Can failing schools be fixed?* Washington, DC: Thomas B. Fordham Foundation.
- The Brown Center. (2003). *Charter schools: Achievement, accountability, and the role of expertise*. Washington, DC: Brookings Institution Press.
- Buchanan, B. (2006). *Turnover at the top: Superintendent vacancies and the urban school*. Lanham, MD: Rowman & Littlefield.
- Bulkley, K., & Wohlstetter, P. (2003). *Taking account of charter schools: What's happened and what's next?* New York: Teachers College Press.
- Carnoy, M., & Loeb, S. (2002). Does external accountability affect student outcomes? A cross-state analysis. *Educational Evaluation and Policy Analysis*, 24, 305–331.
- 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, MA: 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/ericll/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, 249–260.
- Cooper, C. L., & Robertson, I. (Eds.). (1986). *International review of industrial and organizational psychology*. Chichester, UK: Wiley.
- Darling-Hammond, L. (2004). Standards, accountability, and school reform. *Teachers College Record*, 106, 1047–1085.
- Diamond, J. B., & Spillane, J. (2004). High-stakes accountability in urban elementary schools: Challenging or reproducing inequality? *Teachers College Record*, 106, 1145–1176.
- Dorn, S. (1998). The political legacy of school accountability systems. *Education Policy Analysis Archives*, 6(1), Retrieved June 10, 2009, from <http://epaa.asu.edu/epaa/v6n1.html>
- Education Commission of the States. (2004, March). *State takeovers and reconstitutions*. Retrieved March 28, 2006, from <http://ecs.org/clearinghouse/51/67/5167.htm>
- 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.
- 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 schools. *American Educational Research Journal*, 44, 594–629.
- Firestone, W., Schorr, R., & Monfils, L. (Eds.). (2004). *The ambiguity of teaching to the test: Standards, assessment, and educational reform*. Mahwah, NJ: Lawrence Erlbaum.
- Frey, B. S. (1997). On the relationship between intrinsic and extrinsic work motivation. *International Journal of Industrial Organization*, 15, 427–440.
- Fullan, M. (2003). *Change forces with a vengeance*. London: Routledge Farmer.
- Fuller, B., Gesicki, K., Kang, E., & Wright, J. (2006). *Is the No Child Left Behind Act working? The reliability of how states track achievement* (Working Paper No. 06-1). Berkeley, CA: Policy Analysis for California Education.
- Fuller, B., Wright, J., Gesicki, K., & Kang, E. (2007). Gauging growth: How to judge No Child Left Behind? *Educational Researcher*, 36, 268–278.
- Garland, L. (2003). *Navigating treacherous waters: A state takeover handbook*. Lanham, MD: Scarecrow.
- Gill, B., Zimmer, R., Christman, J., & Blanc, S. (2007). *State takeover, school restructuring, private management, and student achievement in Philadelphia*. Santa Monica, CA: RAND.
- Haertel, E. H. (2005). *Using a longitudinal student tracking system to improve the design for public school accountability in California*. Palo Alto, CA: Author.
- Haney, W. M. (2000). The myth of the Texas miracle in education. *Education Policy Analysis Archives*, 8(41). Retrieved August 22, 2006, from <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, and 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, 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 annual meeting of the American Educational Research Association, New York.
- Heistad, D. (2006). *Analysis of 2005 supplemental educational services in Minneapolis Public Schools: An application of matched sample statistical design*. Minneapolis, MN: 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.
- Hess, G. A. (2003). Reconstitution—three years later: Monitoring the effect of sanctions on Chicago high schools. *Education and Urban Society*, 35, 300–327.
- 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.
- Kelley, C., & Protsik, J. (1997). Risk and reward: Perspectives on the implementation of Kentucky's school-based performance award program. *Educational Administration Quarterly*, 33, 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, and 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, MA: The Civil Rights Project at Harvard University.
- Kim, J., & Sunderman, G. L. (2005). Measuring academic proficiency under the No Child Left Behind Act: 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, and school reform* (pp. 9–26). Thousand Oaks, CA: Corwin Press.
- Koretz, D. M., Barron, S. I., Mitchell, K. J., & Stecher, B. M. (1996). *Perceived effects of the Kentucky Instructional Results Information System*. Santa Monica, CA: RAND.
- Latham, G., & Pinder, C. (2005). Work motivation theory and research at the dawn of the twenty-first century. *Annual Review of Psychology*, 56, 485–516.
- Lawler, E. (1973). *Motivation in work organizations*. San Francisco: Jossey-Bass.
- Lee, J. (2006). *Tracking achievement gaps and assessing the impact of NCLB on the gaps: An in-depth look into national and state reading and math outcome trends*. Cambridge, MA: The Civil Rights Project at Harvard University.
- Lee, J. (2007). *The testing gap: Scientific trials of test-driven school accountability systems for excellence and equity*. Charlotte, NC: Information Age.
- Lee, J. (2008). Is test-driven external accountability effective? Synthesizing the evidence from cross-state causal-comparative and correlational studies. *Review of Educational Research*, 78, 608–644.
- 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, 797–832.
- Leithwood, K., Steinbach, R., & Jantzi, D. (2002). School leadership and teachers’ motivation to implement accountability policies. *Educational Administration Quarterly*, 38, 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), Retrieved June 29, 2005, from <http://epaa.asu.edu/epaa/v13n33/>.
- Linn, R. L. (2008). Toward a more effective definition of adequate yearly progress. In G. L. Sunderman (Ed.), *Holding NCLB accountable: Achieving accountability, equity, and 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(9), Retrieved January 19, 2009, from <http://epaa.asu.edu/epaa/v16n9/>
- Lortie, D. C. (1977). *Schoolteacher: A sociological study*. Chicago: 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, 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.
- Mathis, W. J. (2009). *NCLB’s ultimate restructuring alternatives: Do they improve the quality of education?* East Lansing, MI: The Great Lakes Center for Education Research & Practice.
- McCaffrey, D. F., Lockwood, J. R., Koretz, D. M., & Hamilton, L. S. (2003). *Evaluating value-added models for teacher accountability*. Santa Monica, CA: RAND.
- 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: 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. (2008). Low-performing schools’ programs and state capacity requirements: Meeting the NCLB educational goals. In G. L. Sunderman (Ed.), *Holding NCLB accountable: Achieving accountability, equity, and school reform* (pp. 137–151). Thousand Oaks, CA: Corwin Press.
- 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). Retrieved December 3, 2005, from <http://epaa.asu.edu/epaa/v13n48/>
- 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, 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, Orlando, FL.
- Moore-Johnson, S. (2007). *Finders and keepers: Helping new teachers survive and thrive in our schools*. San Francisco: Jossey-Bass.
- Neal, D., & Schanzenbach, D. W. (2007). *Left behind by design: Proficiency counts and test-based accountability*. Cambridge, MA: National Bureau of Economic Research.
- Neild, R. C., & Spiridakis, K. (2002). *Preliminary results: Philadelphia new teacher survey*. Philadelphia, PA: Research for Action.
- Neild, R. C., Useem, B., Travers, E., & Lesnick, J. (2003). *Once and for all: Placing a highly qualified teacher in every Philadelphia classroom*. Philadelphia: Research for Action.
- New York State Education Department, Office of New York City School Improvement and Community Services. (2003). *2002-03 Registration Review Initiative: A summary of the Registration Review report findings*. New York: Author.
- Novak, J. F. B. (2003, December). *Penalizing diverse schools? Similar test scores but different students bring federal sanctions*. Berkeley: 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. (2006). Keeping an eye on state standards. *Education Next*, 6(3), 28–29.
- 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, 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(3), 48–55.
- Reeves, D. (2000). *Accountability in action*. Denver, CO: 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 annual meeting of the American Educational Research Association, New York.
- Saltman, K. J. (2005). *The Edison Schools*. New York: Routledge.
- Scheerens, J., & Bosker, R. (1997). *The foundations of educational effectiveness*. Oxford, UK: Pergamon.
- Sennett, R. (2006). *Culture of new capitalism*. New Haven, CT: Yale University Press.
- Shamir, B. (1990). Calculations, values, and identities: The sources of collectivistic work motivation. *Human Relations*, 43, 313–332.
- Shamir, B. (1991). Meaning, self and motivation in organizations. *Organization Studies*, 12, 405–424.
- Skrla, L., & Scheurich, J. (2004). Displacing deficit thinking in school district leadership. In L. Skrla & J. Scheurich (Eds.), *Educational equity and accountability* (pp. 109–132). New York: Routledge Farmer.
- Spillane, J., Diamond, J., Burch, P., Hallett, T., Jita, L., & Zoltners, J. (2002). Managing in the middle: School leaders and the enactment of accountability policy. *Educational Policy*, 16, 731–762.
- 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, 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.
- Travers, E. (2003a). *Characteristics of schools under diverse providers, 2002–2003*. Philadelphia: Research for Action.
- Travers, E. (2003b). Philadelphia school reform: Historical roots and reflections on the 2002–2003 school year under state takeover: A talk initially presented at the Friends Association of Higher Education and Friends Council on Education Conference at Swarthmore College and Pendle Hill. *Penn GSE Perspectives on Urban Education*, 2(1). Retrieved November 2, 2005, from <http://www.urbanedjournal.org/commentaries/c0007.pdf>
- 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>
- U.S. Department of Education. (2009, January 8). *Secretary Spellings approves additional growth model pilots for 2008–2009 school year*. Retrieved May 4, 2009, from <http://www.ed.gov/news/pressreleases/2009/01/01082009a.html>
- U.S. Government Accountability Office. (2007). *No Child Left Behind Act: Education should clarify guidance and address potential compliance issues for schools in corrective action and restructuring status*. Washington, DC: Author.
- Useem, E. (2005). *Learning from Philadelphia's school reform: What do the research findings show so far?* Philadelphia: Research for Action.
- Valenzuela, A. (2005). *Leaving children behind: How "Texas-style" accountability fails Latino youth*. Albany: State University of New York Press.
- Weiss, M. J. (2008a). The "growth model" pilot isn't what you think it is. *Education Week*, 27(42), 28–29.
- Weiss, M. J. (2008b). *Using a yardstick to measure a meter: Growth, projection, and value-added models in the context of school accountability*. Unpublished doctoral dissertation, University of Pennsylvania, Philadelphia.
- Ziebarth, T. (2002). *State takeovers and reconstitutions* (Rev. ed.). Denver, CO: Education Commission of the States.
- 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.

## AUTHORS

**HEINRICH MINTROP** is an associate professor in the Graduate School of Education at the University of California, Berkeley, 3647 Tolman Hall, Berkeley, CA 94720; [mintrop@berkeley.edu](mailto:mintrop@berkeley.edu). As a researcher, he explores issues of school improvement and accountability in both their academic and civic dimensions.

**GAIL L. SUNDERMAN** is a senior research scientist and director of the Mid-Atlantic Equity Center at the George Washington University Center for Equity and Excellence in Education, 1555 Wilson Boulevard, Suite 515, Arlington, VA 22209; [gsunderman@ceee.gwu.edu](mailto:gsunderman@ceee.gwu.edu). Her research focuses on educational policy and politics and urban school reform, including the development and implementation of education policy and the impact of policy on the educational opportunities of at-risk students.

Manuscript received January 25, 2009

Revision received May 7, 2009

Accepted May 8, 2009