

**Low-performing Schools Programs and State Capacity
Requirements:
Meeting the NCLB Educational Goals**

*Heinrich Mintrop
University of California, Berkeley*

The federal *No Child Behind* Act (NCLB) is an unprecedented national response to the persistent achievement gap in U.S. public schools between children from poor and ethnic minority backgrounds and white middle class children. The mechanisms with which the states are to steer school performance according to the law, however, are not entirely new. A number of “first-generation” states began experimenting with outcome-based accountability systems some ten years prior to the passage of the federal act. Roughly following the NCLB blueprint, these systems were composed of state standards, assessments, a small number of numeric performance goals, support for school improvement, and sanctions for schools (and districts) persistently failing to meet performance targets. But first-generation systems differed widely as to the rigor of their performance demands and scope of low-performing schools programs, scope defined here as the degree of focus, intensity, and comprehensiveness of capacity building in relationship to the acuteness of pressure.

While the federal act imposes some uniformity on the states regarding formal assessment procedures and sanctions, it leaves the definition of educational goals and the actual business of school improvement to the states, allowing design differences to continue. Then, as now, states opted for different levels of rigor and support for school improvement. Thus, the first-generation accountability systems, with their extended performance history and wide variation, provide a useful “laboratory,” from which we

can infer important lessons about NCLB implementation.¹ The focus of this paper is on gauging the scope of low-performing schools programs and the required state capacities to implement them.

The paper draws from the literature on low performing schools, generally, and my own research in this area, most notably the pre-NCLB accountability systems of Maryland and Kentucky, a cross-system analysis of “first generation accountability systems,” and recent research on the California accountability system. Data and reports from three smaller states (Kentucky, Maryland, North Carolina) and two larger ones (California, Texas) are used.² In addition we looked at three urban districts: New York, Chicago, and Philadelphia.

My argument in this paper is as follows:

- Worthwhile goals are the foundation of a good accountability system.³ The need for capacity building in struggling schools is determined by the system’s educational “goal horizon.”
- School improvement within an accountability environment remains an enterprise of muddling through, despite new diagnostics and sanctions, and contrary to the idea of ‘proven’ strategies.
- Reaching educational goals of at least medium rigor requires low-performing schools programs that are sophisticated and intensive.
- Even the most elaborate state programs for low-performing schools to date are not up to the task of successfully reaching NCLB goals. The challenges for state capacity are enormous.

Goal Horizons and the Need for Support

Worthwhile performance demands within outcome-based accountability systems flow from definitions of the good school, and in the context of NCLB from visions of

educating poor children and children of color. There is broad agreement that all children need to learn basic skills, i.e. literacy, numeracy, and for the large number of non-native speakers, ability to communicate in English. And there is a solid body of research on effective schools and effective instruction⁴ that has yielded characteristics of schools and instructional programs from which reformers can usefully draw when they design improvement strategies. The problem is that this body of literature has very little to say about more complex goals and schooling processes; and research about the latter is by far not as unanimous.⁵

The relationship between one's educational goals and values, on one hand, and the dynamics of organizational reform one is likely to embrace, on the other hand, seems to be bi-directional. While in earlier historical periods educational goals or values may have driven organizational change dynamics (as in the school wars of the sixties⁶), the current phase of high-stakes accountability rests on an inversion of this relationship. It does not seem to be fueled as much by political, ideological or moral zeal about aims, but foremost by certitudes about rational principles of organizational development and productivity that are borrowed from the world of business. Thus, in our times, ideas about *what works* drive the educational aims and notions of quality that come into prominent view.

Many proponents of outcome-based accountability systems, the early designs of the Maryland and Kentucky systems being a case in point, assume that complex educational goals can be reached within this model. And there is a priori no reason why they may not, given proper design. But it seems that, in light of the evidence discussed in this paper, the specific architecture of NCLB may set us up for a rather limited horizon.

One could imagine an accountability system that hones in on the most basic of skills, poses within-easy-reach performance targets, puts high pressure on reaching these targets, and measures them with repeatedly-used test forms. In all likelihood, we may have designed something that works, probably requiring little capacity building investment, but do we value it, given the propensity in such a system for test-training rather than substantive learning? And so, when we reflect on designs for low-performing schools programs (and states' reform capacities to implement them), we need to persistently ask ourselves, not only if the programs work, but also if we value what they deliver.

Michael Fullan drew the following lesson from his extensive evaluation of national systemic reform efforts in English schools. He found that in the first years schools in the system improved remarkably, but noted that this improvement came with a price: on the side of the teachers, serious morale problems, a high degree of program prescriptiveness, and teacher dependency; on the student side, a stagnation of performance once the low-hanging fruit, that is organizational tightening up and improvements on basic skills, had been harvested. In Fullan's language this is Horizon Number 1.⁷

No question, many schools would greatly benefit from improvements in those areas, but it is troublesome if a system as a whole gets stuck on this level. Low-performing schools programs need to encourage schools to move towards Horizon Number 2, characterized by reasoning and problem solving, while not losing sight of teaching the basic skills. If we just look at what works within Horizon Number 1, we have a model that may be workable within the present architecture of NCLB, given

current efforts at school capacity building: Fairly low test rigor, goals pegged to presently available state and teacher capacity, pressure on districts and schools, and consequently fairly light, i.e. affordable and manageable capacity building needs.⁸

Conceivably, states could phase in the system with a heavy emphasis on Horizon 1 and gradually increase the cognitive complexity of its performance demands. It is possible that such a gradual and cautious approach, over a generation or two, may turn out to be effective. If states wanted to overcome schools' and communities' low expectations and prove, in a first step, that improvements are indeed possible in the lowest performing schools, then such a gradualist design may be advantageous compared to systems that reinforce a sense of failure with demands that are over ambitious relative to available capacities. But there are two problems with the gradual approach: (1) it will most likely not close the achievement gap beyond the simple skills of Horizon 1, if that, within the short time frame of NCLB; and (2) if the system relies heavily on pressure, it may squelch teachers' initiatives in Horizon 2, with serious repercussions for educational quality down the road.

NCLB leaves it up to the states to define test rigor and proficiency levels, though the federal NAEP tests (National Assessment of Educational Progress) function as a one benchmark for gauging the rigor of state assessments.⁹ Tables 1 and 2 show how testing rigor fundamentally structures a state's intervention burden in low-performing schools.

Table 1. Differences between Percent of Students Scoring Proficient on NAEP and State Tests (2003).

	CALIFORNIA			TEXAS			KENTUCKY		
	NAEP	State	Difference	NAEP	State	Difference	NAEP	State	Difference
4 th grade Reading	21	39	18	27	85	58	31	62	31
8 th grade Reading	22	30	8	26	88	62	34	57	23
4 th grade Math	25	45	20	33	87	54	22	38	16
8 th grade Math	22	30	8	25	72	47	24	31	7

Source: Education Week, vol. 24, no. 17, (Jan. 6, 2005)

Table 2. Numbers of Schools in Need of Improvement Based on AYP (2003-04).

	CALIFORNIA	TEXAS	KENTUCKY
<i>N</i>	1,626	199	130
Percent	~20%	~6%	~12%

Source: Education Week, vol. 24, no. 17, (Jan. 6, 2005)

States, such as California, with high testing rigor in combination with challenging demographic conditions produce an enormous intervention burden while states with less rigorous tests and more lenient definitions of proficiency, such as Texas, face a relatively modest challenge. Kentucky is a state with medium testing rigor and correspondingly a medium intervention burden. Thus, if states were to adopt definitions of proficiency close to NAEP, as California did, the result would be a high number of schools in need of school improvement.

School Improvement in Outcome-based Accountability Systems

School improvement is an intricate business. Whether a school succeeds to improve is dependent on a host of factors. Factors, internal and external to the organization, come into play. The motivation and capacity of the workforce, the strength of interactions among staff, the school's programs for students' cognitive, emotional and social development, and the implementation of specific improvement strategies are to a large degree under the control of schools. The supply of material and human resources, the design of sound policies, regulations, and incentives, the adoption of effective programs, and the provision of technical assistance are to a larger degree externally generated, primarily by districts and states, but also by non-governmental third-party providers. School life is fundamentally situated in the socio-economic status and culture of the community the school serves and in an educational market competition for students. As these multiple factors interact with each other, they produce idiosyncratic constellations that make success in school improvement efforts an uncertain and contingent outcome for individual schools.

Outcome-based accountability systems are designed to introduce a greater degree of rationality into school improvement. Most notably, they reduce the number of goals to a few easily measurable achievement targets and set up a streamlined incentive system by adding pressure to reaching these targets. Given higher goal clarity and more performance information, they encourage the development and adoption of purportedly robust programs and school restructuring efforts that work. Clear goals, focused pressures and ready-made programs and interventions seem to simplify the task of school improvement. The record from first-generation accountability systems, however, cautions us not to overestimate these rationality claims.

Sanctions

Pressure and the threat of more severe sanctions were a conspicuous feature of low-performing schools programs when high-stakes accountability systems first came into existence in the 1990s, and they are a prominent feature in NCLB. Practically all of the sanctions suggested by NCLB had been on the books or been tried by the first-generation systems examined here. State takeover was the most severe sanction in the Maryland system.¹⁰ Public hearings, appointment of a special on-site monitor or master, and eventual school closure were envisaged by the Texas regulations as sanctions.¹¹ Assignment of an instructional officer, external partner, removal of the principal, and school reconstitution figured prominently in the Chicago system.¹² Redesign and closure were also primary sanctions in the New York Schools Under Registration Review (SURR) program.¹³ Kentucky and North Carolina added penalties to this list that touch individual teachers more severely.¹⁴ But these sanctions were very rarely imposed and their centrality faded over time. Kentucky is a good example. The original language of schools “in decline” and “in crisis” was replaced by schools “in need of assistance.”¹⁵ The state-appointed “Distinguished Educators,” who initially combined technical assistance and probation management in their role, were renamed “Highly Skilled Educators” and shed their evaluative function.¹⁶ Actual imposition of final sanctions has been a negligible feature in Kentucky.¹⁷

In Texas, more severe sanctions akin to the level of corrective action were used very sparingly.¹⁸ In 2002, there were seven schools under the supervision of a monitor who has little authority, and two schools under the supervision of a master who has authority over the local district.¹⁹ The state reconstituted only a handful of schools.²⁰

Texas primarily relied on the threat of bad publicity to motivate districts and schools to improve performance.²¹ Likewise in Maryland, after five years of high stakes accountability, the state finally took over four schools and assigned them to private management organizations.²²

When the present California accountability system was designed with a plethora of NCLB-type sanctions,²³ the turn from pressure to support that earlier accountability systems seemed to have undergone was evident. The California program already began with voluntary participation of qualifying schools.²⁴ Schools selected into the program accepted increased scrutiny and accountability from the state in return for funds usable for capacity building at the site.²⁵ When fewer schools than envisioned met their growth targets, the state refrained from building up pressure. It readjusted growth expectations and added additional intervention layers preceding more severe sanctions.²⁶

Why this turn from pressure to support? Some suspect that states shrink from the responsibility and political costs that the heavy hand of sanctions entails.²⁷ This is one plausible explanation, but other research suggests that, political costs notwithstanding, the pressure strategy is not as promising a motivation strategy as perhaps originally perceived.²⁸ This is so for a number of reasons. Increasing pressure on schools that do not have the baseline capacity to meet performance goals is not motivating; rather it further fragments already precarious organizations. Sanctions turn off high performing teachers that are present in most low-performing schools. And lastly, sanctions, widely seen as unfair for schools that educate poor children, violate the professional norms, values, and expectations of many teachers.

Thus, in their majority, first generation states have either rarely used, or turned away from, high pressure 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. If experiences of the first-generation accountability systems are any indication, states are advised not to rely too much on the power of pressures and sanctions to bring about needed improvements.

Proven Strategies

School improvement would indeed be rather simple if it merely was about matching an identified performance problem with a proven strategy implemented by willing educators. The record from first-generation systems speaks otherwise. We already saw the limitations of a motivational strategy that banks on pressure. But there are also problems with the idea of proven strategies. A number of these strategies have been tried for corrective action and school redesign within first-generation accountability systems. They seemed to have worked in some contexts, but not in others,²⁹ confirming the contingent nature of school improvement even within the context of stringent accountability systems. I will briefly summarize findings on the most commonly used strategies.³⁰

Reconstitution. 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 number of schools also benefited from the fresh start.³¹ Results from Chicago's reconstitutions were inconclusive as well.³²

Fundamentally, staff replacements were not necessarily of higher quality than the original teaching staff, and in many schools teacher morale plummeted.³³ In New York's SURR program corrective action and redesign were used more vigorously. Almost fifty schools were reconstituted.³⁴ More than a tenth of the schools were closed.³⁵ Some schools benefited, yet only about half (153) of the SURR schools have exited the program successfully so far.³⁶

Educational Management Organizations. Maryland took over four schools from the Baltimore City school district and passed them on to two educational management organizations (EMO).³⁷ Under one of the EMO's, only one of its three schools saw consistent gains, one performed unevenly, and one was not improving.³⁸ In Philadelphia, we have higher numbers of schools that were taken over. One fourth of all district schools were taken over, with 45 managed by different external management organizations and 25 by the district's newly created Office of Restructured Schools. Here, each provider offers different models of intervention.³⁹ Preliminary data, at the time of this research, suggest that takeover by management companies has helped in some cases, but is not universally positive.⁴⁰

External Partners. This feature was widely used in Chicago where each school on probation was assigned an external partner⁴¹. Originally, external partners developed their own models of intervention, but disparities in the quality of services concerned the district.⁴² An inherent problem in external partner, as well as diverse provider, models is the lack of focus on state or district goals and the uneven quality of provided consultant services.⁴³

Charters. 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.⁴⁴ Available data seem to suggest that converting district-administered schools into charter schools has had uneven results. A multi-state study by the Brown Center on American Education shows that generally charter schools lag behind, or are similar to, regular public schools in absolute performance and gains from year to year.⁴⁵ Charter schools also tend to show up on states' lists of failing schools in larger proportions than regular public schools. Anecdotal evidence from Philadelphia suggests that charter school conversion without the benefit of an external provider model may be the least successful conversion of the ones tried there.⁴⁶

District Takeovers. State takeovers of entire districts have also produced uneven outcomes. Financial management is often cited as the most promising area for potential success by states.⁴⁷ However, equally dramatic *academic* success has been much harder to achieve.⁴⁸ Academic gains have been mixed at best, most often occurring only after multiple years of intervention.⁴⁹

Intervention Teams. These are teams that enter schools as authoritative interveners. They are charged to evaluate schools, prescribe remedies, and help with implementation. In North Carolina, these teams were said to be rather successful; in California they worked with mixed success, encountering much resistance at the school level.⁵⁰ The two states differ with regard to both operational principles and context. The North Carolina teams were recruited by the state from the ranks of seasoned practitioners and closely worked with schools on an almost daily basis.⁵¹ As teachers in North Carolina cannot engage in collective bargaining, teacher unions are less of a force.⁵² In California, the

teams were either third-party providers or county offices of education that traditionally were not involved in the day-to-day affairs of regular district schools.⁵³

In summary, a variety of corrective action strategies have been tried by the examined systems, but none stick 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.

Thus, first-generation accountability systems demonstrate that firstly creating motivation through pressure is not as powerful an option on the ground as it might appear in the language of the law and, secondly, that robust and universally effective interventions are hard to come by. School improvement, even under conditions of stringent accountability, is (and remains) far more complex than matching an identified performance problem with a proven strategy implemented by willing educators. As a consequence, states are advised to design low-performing schools programs that are rich in capacity building and sophisticated enough to address the complexity and contingent nature of the task.

The Need for Capacity Building

By looking at variations among first-generation systems across states, we gain a better understanding of the requirements for capacity building. In two of the four states examined here, Maryland and Texas, capacity building was not a prominent feature of the low-performing schools programs; in two others, California and Kentucky, it was.

Maryland. The toughest challenge ahead was created by the Maryland system in the 1990's, a system that operated, to speak with Fullan, in Horizon 2. The system targeted extremely hard cases in decline, demanded of schools to adjust to highly complex assessments (which fewer than half the state's student population managed to pass with satisfaction), and set the exit criteria for its low-performing schools program very high. The state did not develop an elaborate capacity building structure. State monitors were the eyes and ears of the state. Their role in internal school improvement efforts was minimal. Very few low-performing schools managed to exit the program; and indeed schools statewide stagnated until the system was abandoned. In the Maryland case, state performance demands, pegged to goal Horizon 2, were decoupled from existing capacities, and with a lack of compensatory capacity building pressures became ineffective or counterproductive.

Texas. Texas took an approach that contrasted with that of Maryland in testing rigor, but exhibited similarities regarding capacity building. Operating within Horizon 1, the state pegged performance demands at levels that challenged schools in the bottom 20 to 40 percent of the performance distribution with cognitively simple tests. The state has a decentralized form of governing schools and did not take a strong leadership role in providing support to ailing schools, relative to more intensive efforts in other states.⁵⁴ However, the state required low-performing schools and districts to compile a school improvement plan. It sent peer review teams to schools and districts that visited a school or district for varying lengths of time depending on size of school or district. These peer review teams were made up of state department staff and evaluators that received training

with the help of a CD. In addition the state organized educational support centers that offered their services to low-performing schools and districts, but not exclusively so. Other schools in need of support could contact these centers as well. Texas did not furnish additional monetary grants to low-performing schools. The state, however, had strong mechanisms built into its accountability system that identified low-performing districts directly and threatened them with further sanctions. Given that performance demands were more closely pegged to existing teacher capacities, the state could bank on a pressure strategy that succeeded by motivating schools to harvest the low-hanging fruit,⁵⁵ while keeping the need for support relatively limited. Test scores on the state tests rose and low-performing schools exited the program in large numbers.

California. As we saw earlier, California's rigorous performance demands, relative to NAEP definitions, coupled with challenging student demographics, led to a burgeoning number of identified low-performing schools that experienced mild accountability pressures. For capacity building, the state banked on a massive disbursement of grants attached to a very loosely constructed oversight structure.⁵⁶ Not all schools that could have received these grants did so. The state selected schools according to priority and by chance.

Identified schools had to contract with an external evaluator who was chosen from a state-approved list. Educational reform projects, consultants, county offices of education and later even district offices themselves could apply to this list. Training in evaluation was not provided. The state, however, did require vendors to reapply to the list showing evidence of success. To receive grant money, schools were to write a school improvement plan that was at first given a cursory review by the state department.

Subsequently, this requirement was reduced to a short summary of the plan, the full plan being kept on file locally. Thus, in the California case, the state department kept a low profile. It relied primarily on grant making at a magnitude far greater than most other states we examined, on the capacity of local vendors, the willingness of local districts, and the wisdom of schools to spend the money wisely. A management structure that could ensure quality of the support system was only weakly developed. Reports showed that schools' responses to the program varied widely and depended on the varying quality of external evaluators.⁵⁷ A systematic evaluation of the program⁵⁸ did not show significant program effects. Qualitative data suggest that the schools lacked sustained quality support. The number of low-performing schools in the program remained high.

Kentucky. Of the state programs we surveyed, Kentucky had a fairly elaborate system in place that provided oversight and support to schools under direct supervision from the state department. Services were sustained over one school year or longer, and specifically targeted to low-performing schools achieving state goals. As part of the state's support for its schools "in need of assistance," Kentucky provided modest additional school improvement funds. In the 2002-2003 year, \$2 million was budgeted for 90 schools. For example, elementary school grants ranged from \$12,000-\$38,000 per biennium. A school inspection was conducted by state-sponsored Scholastic Audit Teams, which included a highly skilled educator (HSE), a teacher, a principal or other administrator, a parent, and university-based educator.⁵⁹ The audit teams were trained for their task and visited each school for about a week. Once the scholastic audit was conducted, schools used the results to write their school improvement plans.

HSE's had to demonstrate prior ability to bring about high levels of student performance and went through a rigorous hiring and training process. Each HSE received two weeks of training and follow-up training at quarterly meetings. Mentors from the state department provided assistance in problem solving and support to HSE's. HSE's were expected to serve on-site at least eighty percent of their work time. Their activities included but were not limited to: staff development, classroom observations of instruction, demonstration lessons, grants writing, tutoring, and creation of model lessons.⁶⁰ In addition, a team of HSE's that specialized on organizational management was formed and could be assigned to more than one school at a time, given the needs of particular schools. In the 2002-2003 school year, there were 52 HSE's working with 30 very low-performing schools and providing support to others on a voluntary basis. The program succeeded in exiting most of the schools. Significant challenges for the program were sustaining the change once HSE's had left school grounds, creating an appropriate match between the HSE and the school, and maintaining a strong pool of HSE's.⁶¹

We can infer from the comparison across states that the need for strong state support grows in proportion to performance demands. Programs can be successful without intensive capacity building for struggling schools as long as they operate within a low goal horizon. But as soon as goals move beyond the most basic of skills, the lack of capacity building seems to bode ill for success. California's program disbursed generous grants for capacity building and its decentralized structure made it adaptable to local conditions, but it lacked a management structure at the state level that was strong enough to assure quality of services and focus. Its effect dissipated. Programs, such as the one in Kentucky,⁶² that coupled rigor with focused, intense, and comprehensive support were

fairly successful (as indicated by schools exiting the program). Kentucky's program stresses support over sanctions, supervises this support centrally, and manages recruitment and training of personnel and quality control of services. Services are geared toward the comprehensive reform of schools with a focus on the state's programmatic mandates and performance goals. But at the same time, on-site support providers are sophisticated enough to flexibly adapt their intervention to individual school needs, though curriculum and instructional alignment are key points of intervention.

The necessity for comprehensive support is underscored by New York's Chancellors District, now defunct. This effort, emulated by other inner-city districts, consisted of the following elements:⁶³

- Reduced class size
- Extended school day and year
- After-school program
- Prescribed instructional program, schedule and curriculum
- Professional development: A minimum of four on-site staff developers and a teacher specialist assigned to each school
- Student assessments
- Supervisory and instructional support
- Restaffing and replacement of most principals and many ineffective teachers
- More intense monitoring and mentoring
- Incentives for recruiting qualified teachers (e.g., signing bonuses).

Interviewed researchers and program administrators point to two factors that in their minds made a key difference: the special district removed a school from a failing district and put it in a very nurturing one, and a set of interventions and strategies were given to schools as a bundle, avoiding isolated quick fixes.⁶⁴ However, even with this intense intervention, data suggest that Chancellor's District schools achieved only moderate improvement in student performance; only half of the enrolled schools were removed

from the state list of low-performing schools, and one-fifth had to be closed.⁶⁵ A sobering result that clarifies the pervasive need for capacity building in the most challenging educational environments.

NCLB Goals and Program Scale

The goals of NCLB are ambitious and commendable: rigorous levels of proficiency should be reached by *all* students within a fairly short time frame. Rigorous definitions of proficiency, for example by using NAEP criteria, would push states' performance demands well into goal Horizon 2 – and would in all likelihood create a huge intervention burden. The California system with a proficiency definition fairly close to NAEP and an intervention burden of a fifth to a quarter of all schools would be a likely scenario for the country as a whole (or at least for states with similarly large numbers of poor and immigrant students). That this scenario has not become reality is due to the fact that many states eschew these high standards and the specter of unmanageable loads of troubled schools, in need of support and at some point requiring sanctions.

Echoing this concern, first-generation accountability systems typically kept their programs on a manageable scale or down-scaled them over time. In Maryland, a state with a relatively large program, the state department limited the burden by capping the number of schools at around a hundred (about 7% of all schools) although many more schools could have qualified according to the state's criteria. Texas kept the intervention burden even smaller. In 1995, the system identified 267 low-performing schools. The numbers dropped to 59 in 1998, and rose again continuously to 150 in 2002.⁶⁶ The

thresholds for entrance and exit had risen in the meantime, but the state department saw to it that state capacity was not overwhelmed. With these numbers, the program fluctuated in the 2 to 4% range of the total number of schools in the state. The Kentucky program started out with 250 schools, or about 20 percent of all schools. But these numbers were swiftly curtailed. In the 2002 accountability cycle, the state identified merely 90 schools as low performing or about 7.5% of the total.⁶⁷ Only one third of those were required to accept state intervention which in Kentucky's case, as we saw, was intensive.⁶⁸

If the record of first-generation systems is any indication, it seems that over time state intervention burdens leveled off at around to 2 to 4 percent of the total number of schools. Even in California, with large numbers of identified schools, the state severely curtailed the number of schools that received grant money from the state. Moreover, with educational goals of at least medium rigor, even a small intervention burden requires a sophisticated, intensive, and comprehensive capacity building effort. But even comprehensive approaches, as we saw in the case of the special district in New York City, are sometimes hard-pressed to overcome performance barriers that exist in the highest-need and lowest-capacity schools and districts.

One can infer from the records of the most developed first-generation accountability efforts that a strategy of realistically meeting the ambitious NCLB goals, i.e. closing the achievement gap based on rigorous educational goals in a fairly short time period, would require low-performing schools programs on an unprecedented scale with an intensity, focus, and comprehensiveness of capacity building that has heretofore only been tried for relatively small numbers of schools. The alternative would be either a capping of the

number of identified schools tailored to available state capacity, a strategy that a number of states pursued, but is no longer an option under NCLB, or a down-scaling of performance demands while increasing pressure, confining schools to a low goal horizon.

Conclusion

The purpose of this paper is to gauge the scope of low-performing schools programs needed to reach ambitious NCLB goals and the level of state capacity required to implement these programs. I draw my conclusions from the literature on low-performing schools and especially the successes and shortfalls of first generation accountability systems that preceded NCLB by up to a decade.

First and foremost, we need to design systems that work, but at the same time deliver on educational goals we value. An accountability system that emphasizes basic skills, poses within-easy-reach performance targets, i.e. within the margins of available educator capacity, and uses high pressure to reach these targets may “work” without expanding investments in school and state capacity. But if such systems constrain the education of poor children within, what Fullan calls, goal Horizon 1, do we value these systems enough to advocate for them?

As soon as one moves into the realm of at least medium rigor, pressures and sanctions have shown to be a rather blunt instrument. Corrective action strategies, enumerated in NCLB and tried in first-generation accountability systems, have shown to be subject to the usual contingencies of school improvement of which I spoke earlier. Thus, severe sanctions do not work as a fallback solution, and fresh-start measures heighten the need for sophisticated, context sensitive support that can counteract the

usual school improvement contingencies. Otherwise the promise of a fresh start may get lost in unproductive turbulence.

After an initial ‘high-flying’ phase, even the most vigorous and stringent pre-NCLB accountability systems restricted their intervention load to below ten percent of total number of schools. Even this relatively reduced scale required focused, intensive, and comprehensive support for struggling schools, the scope of which grew with the increasing educational rigor of the system. States used various designs to meet schools’ needs. Small states organized a more centralized effort of support provision, while larger states either relied on regional centers or third-party providers. But independent of size, performance demands of at least medium rigor require elaborate capacity building structures. Whether support and oversight is provided directly by the state or through third-party consultants, low-performing schools programs need a management structure that allows for careful recruitment and quality control of service providers.

Contrast this with the reform dynamic implied in NCLB: intervention loads attuned to NAEP proficiency definitions (in many states probably up to a third of all schools over time), strong emphasis on sanctions, and the idea that robust, research-based models of improvement are readily available. It is in the logic of a federal system that uniformity in assessment procedures and sanctions begs for more uniformity in the substance of acceptable performance; and the featuring of NAEP is a further nudge in this direction. First-generation systems suggest that NAEP-adequate rigor would result in intervention loads, capacity building needs, and required state administrative capacity on an unprecedented and untried scale.

First-generation attempts have shown that the task of continuous school improvement requires states and districts to ‘move on all fronts’ and go beyond incentives and sanctions. Even generous additional grants for capacity building are not sufficient. The enormity of the task at hand requires the federal government, states, districts, and schools to search for powerful, high quality and comprehensive ways of reform and institution rebuilding. Alternatively, states could reduce testing rigor or keep rigor down; and we could be faced with the undesirable trade-off between the ends of achieving basic literacy and numeracy by means of severely curtailing the spectrum of educational goals. In Lauren Resnick’s words,⁶⁹ we would have succeeded in creating a 21st century accountability system that delivers on a 19th century model of learning.

Recommendations

Goal Setting:

Unrealistically ambitious goals relative to schools’ available performance capacity and states’ support for further capacity building make an accountability system dysfunctional. In the long run, such a system becomes illegitimate and de-motivating. The current NCLB architecture seems to encourage systems to operate in the low-rigor range with undesirable consequences for the education of children from poor families. To remedy this situation, the law ought to stipulate state proficiency definitions of at least medium rigor as indicated by NAEP for all states. Rather than setting a fixed time line for all states reaching proficiency, growth goals for each state ought to be pegged to real growth achieved by a sizable number of demographically similar Title I schools in a given state. This ensures that high performing schools are not merely exceptional outliers. As a consequence, lower performing schools have realistic goals to aim for. In order to

avoid system overload, growth goals should be gauged so that no more than 10 percent of the total number of schools are identified as low-performing and in need of support at any given time.

Rather than mandating fixed goals, the federal government should encourage accountability system design competitions among states. These design competitions should be facilitated by an independent and authoritative non-partisan commission or non-governmental agency that monitors state progress based on multiple indicators: differential growth on state tests and NAEP for various levels of cognitive complexity, completion rates, years-to-completion, etc. The agency's effort ought to be supported by independent researchers that are recognized in their field as specialists. Monitoring of multiple data should result in yearly ratings of states' progress. Exceptionally strong designs are then publicly and authoritatively recognized. Thus, states have a strong incentive to learn from powerful designs, but are not regulated as to the details of their approaches.

Sanctions:

Over reliance on sanctions leads to undesirable distortions, such as restrictive goals within Horizon 1 and a narrowing of taught curriculum to measured indicators and accountable students. Unduly high pressure leads in many instances to hasty quick fixes, the implementation of context-insensitive "solutions," and to organizational fragmentation rather than organizational health. Inflexible staging and limited intervention menus lead to unproductive turbulence, rather than a fresh start and sustainable renewal. In order to avoid this situation, the law should stipulate that districts

develop, based on state growth goals, a growth plan for each identified low-performing school. In this plan, schools and districts commit to realistic and complex goals and multi-year programmatic improvement. Ordinarily, districts hold schools accountable to this plan through regular oversight and lines of authority. States give advice on, approve, and monitor individual school plans or district-wide improvement plans. The latter applies if high numbers of schools in a given district are identified as low performing. For the district, reaching state performance goals is high-stakes, involving increasing state oversight and loss of governance autonomy. Districts, in cooperation with state agencies, decide what mixture of pressure, sanctions, corrective action, or support they want to exert on their low-performing schools, depending on local conditions and available alternative resources. The state empowers districts to circumvent local agreements if necessary.

Support Structures:

Reaching performance goals of at least medium rigor in all schools hinges on sizable state support for low-performing schools and districts. Given the enormous variation in states' size, political culture, administrative structure, degree of centralization and educational reform history, it does not make sense to mandate a specific program design for all states. But a number of minimal features ought to be stipulated for state accountability and support systems:

- Multiple performance indicators capture the complexity of educating children and reflect professional norms of good education. These indicators could be student attitudes (e.g., discipline, engagement, academic challenge, safety, etc.); tests of basic skills and problem solving; performance tasks.
- Multiple process indicators help states, districts and schools to steer and evaluate the improvement process. These indicators could be: strength of leadership; faculty cohesion; openness to learning; instructional coherence and creativity, teacher commitment to stay, etc.
- In a first round, schools are *identified* as low-performing if they fail to meet state growth goals. In a second round, multiple performance and process indicators are taken into consideration when states *classify* schools as low-performing. State audits that can assess a school's organizational health with sophisticated professional judgment intervene between identification and classification. Such audits help the accountability system to become better anchored in educators' sense of system fairness and validity.
- Program essentials should be formulated that the state commits to implement in all low-performing schools. Essential components could be: Core academic programs and instructional materials; sufficient instructional time in all subjects; enrichment in sports and the arts; professional development for teachers and principals; curriculum-embedded benchmark monitoring; on-site instructional reform assistance; team collaboration time; improvement grants.
- Capacity essentials should provide for adequate facilities; structures for student well-being (e.g., counselors, deans, personalized attention in small classes, advisories); minimal teacher qualifications; structures for teacher development and collective decision making (e.g., planning time, leadership teams).

- School intervention should target curricular and organizational issues. Services could be organized in state teams, single change agents, district-maintained reform facilitators, non-governmental service providers, consultants, management companies, universities, etc. It is up to states to design a comprehensive plan subject to federal approval.
- State audit and intervention personnel should have expertise in fiscal, organizational, and instructional areas so that they are able to cooperate with districts in designing and reviewing individual school plans and to intervene in district affairs when needed.
- At minimum, a central state management structure sets central goals, selects support providers, organizes the training, and oversees quality control. Services need to be of professional caliber, non-bureaucratic and not merely compliance oriented. States may opt for more centrally developed solutions for system-wide problems.
- An intervention burden of up to 10 percent of total number of schools seems a realistic frame for funding needs with medium rigor performance demands. The law should make a fiscal commitment to fund these capacity building needs.

¹ For more on these different approaches, see Zena Rudo, *Corrective Action in Low-Performing Schools and School Districts* (Austin, Tx: Southwest Educational Development Laboratory, 2001).

² H. Mintrop, *Schools on Probation: How Accountability Works (And Doesn't Work)* (New York: Teachers College Press, 2004); H. Mintrop and T. Trujillo, "Corrective Action in Low-Performing Schools: Lessons for NCLB Implementation from State and District Strategies in First-Generation Accountability Systems." *Educational Policy Analysis Archives* 13, no. 48 (2006). The comparative analysis of evidence across states in this paper summarizes the findings from the earlier report. I wish to acknowledge Tina Trujillo's contribution to this report.

³ A. Porter and M. Chester, "Building a High-Quality Assessment and Accountability Program: The Philadelphia Example" *Brookings Papers on Education Policy* - 2002, pp. 285-337

⁴ J. Scheerens, & R. Bosker, *The Foundations of Educational Effectiveness*. New York: Elsevier (1997). C. Teddlie & D. Reynolds (Eds.), *The International Handbook of School Effectiveness Research*. New York: Falmer (2000).

⁵ B. Creemers, *The Effective Classroom*. New York: Cassell. (1994)

⁶ D. Ravitch, *The Great School Wars*. New York: Basic Books (1974).

⁷ M. Fullan, *Change Forces with a Vengeance*. London: RoutledgeFalmer, (2003). Heinrich Mintrop, "The Limits of Sanctions in Low-Performing Schools," *Education Policy Analysis Archives* 11, no. 3 (2003), <http://epaa.asu.edu/epaa/v11n3.html>

⁸ Paul Paterson and Frederick Hess, *A Race to the Bottom. Keeping on Eye on State Standards*. Retrieved at http://www.educationnext.org/20063/pdf/ednext20063_28.pdf

⁹ It should be noted that in the public discussion, NAEP and state tests are often compared when in many instances they may measure different types of performances. The NAEP benchmark, useful as it is, should therefore be treated with caution.

¹⁰ Maryland State Department of Education, "Fact Sheet #5: School Reconstitution: State Intervention Procedures For Schools Not Progressing Toward State Standards" (2001).

¹¹ *Tx. Ed. Code*. § 39.131.

¹² G. Alfred Hess, "Reconstitution Three Years Later," *Education and Urban Society* 35, no. 4 (2003): 300-327.

¹³ SURR: Schools Under Registration Review. See: Carol Ascher, Norm Fruchter, Ken Ikeda, Schools in Context: Final Report to the New York State Education Department. Analysis of 1997-1998 SURR Schools and Their Districts (New York: Institute for Education and Social Policy, School of Education, New York University, 1999); Ronald Brady, *Can Failing Schools Be Fixed?* (Washington, D.C.: The Thomas Fordham Foundation, 2003); New York State Education Department, Office of New York City School Improvement and Community Services, "The Registration Review Process" (2002).

¹⁴ For a description of Kentucky's accountability system, see Jane David, Pamela Coe, and Patricia Kannapel, *Improving Low-Performing Schools: A Study of Kentucky's Highly Skilled Educator program* (Lexington, Ky: Partnership for Kentucky Schools, 2003); and David Holdzkom, *Low-Performing Schools: So You've Identified Them—Now What?* (Charleston, W. Va.: AEL, Inc., December 2001). For a description of North Carolina's accountability system, see SERVE, *Assisting Low-Performing Schools in the Southeast: A SERVE Special Report* (Greensboro, NC: SERVE, Inc., December 2001); Holdzkom, *Low-Performing Schools*; Helen Ladd and Arnaldo Zelli, *School-Based Accountability in North Carolina: The Responses of School Principals*, Working Papers Series SAN01-13 (Durham, NC: Terry Sanford Institute of Public Policy at Duke, July 2001). For a description of the evolution of North Carolina's teacher testing system, see Kathleen Kennedy Manzo, "N.C. Lawmakers Revoke Teacher-Testing Plan," *Education Week*, June 17, 1998. For a description of the progression of Kentucky's teacher accountability system, see David et al., *Improving Low-Performing Schools*.

¹⁵ David et al., *Improving Low-Performing Schools*.

¹⁶ Ibid.

¹⁷ Interview with Kentucky Department of Education official, December 2003.

¹⁸ Interview with Texas Education Agency official, December 2003.

¹⁹ Texas Education Agency, "2002 Comprehensive Annual Report on Texas Public Schools: A Report to the 78th Legislature," 2002.

²⁰ Interview, Texas Education Agency official; Chris Ferguson, *The Progress of Education in Texas* (Austin, Tx: Southwest Educational Development Laboratory, November, 2000); Charles A. Dana Center researcher, email message to author's research assistant, December 12, 2003.

²¹ Lance Izumi and Williamson Evers, "State Accountability Systems," in *School Accountability: An Assessment by the Koret Task Force on K-12 Education*, ed. Williamson Evers and Herbert Walberg, 198 (Stanford, Ca: Hoover Institute Press Publication No. 512: 2002); Linda Skrla et al., "Accountability for Equity: Can State Policy Leverage Social Justice?" in *Educational Equity and Accountability: Paradigms, Policies, and Politics*, ed. Linda Skrla and James Scheurich, 65 (London: RoutledgeFalmer, 2004); Interview, Texas Education Agency official; and Interview, Texas A & M University professor, December, 2003.

²² Maryland State Department of Education, "Fact Sheet #5;" Council of Chief State School Officers, *State Support*.

²³ *Ca. Ed. Code*. § 52055.5.

²⁴ Jennifer O'Day and Catherine Bitter, *Evaluation Study of the Immediate Intervention/Underperforming Schools Program and the High Achieving/Improving Schools Program of the Public Schools Accountability Act of 1999* (Sacramento, Ca: American Institutes of Research, 2003).

²⁵ Sherry Posnick-Goodwin, "Under Close Scrutiny," *California Educator*, September 2003.

²⁶ O'Day and Bitter, *Evaluation Study*. See also: Interview with WestEd director, December 2003.

²⁷ Brady, *Failing Schools*; Interview, WestEd director.

²⁸ Richard Elmore, *School Reform from the Inside Out*. Cambridge, MA: Harvard Education Press (2006). Heinrich Mintrop, "The Limits of Sanctions in Low-Performing Schools," *Education Policy Analysis Archives* 11, no. 3 (2003), <http://epaa.asu.edu/epaa/v11n3.html>; Heinrich Mintrop, *Schools on Probation: How Accountability Works (And Doesn't Work)* (New York: Teachers College Press, 2004); Consortium for Policy Research in Education (CPRE), *U.S. Department of Education Regional Forum on Turning Around Low Performing Schools: Implications for Policy*, Policy Bulletin pb-01-01 (Philadelphia, Pa: Consortium for Policy Research in Education, 2001); O'Day and Bitter, *Evaluation Study*; Betty Malen et al., "Reconstituting Schools: 'Testing' the 'Theory of Action'," *Education Evaluation and Policy Analysis*, 24, no. 2 (2002): 113-132.

²⁹ Brady, *Failing Schools*.

³⁰ See Mintrop and Trujillo, "Corrective Action..." for details.

³¹ Malen et al., "Reconstituting Schools."

-
- ³² Hess, "Reconstitution."
- ³³ Ibid.
- ³⁴ Education Commission of the States (ECS), *State Interventions in Low Performing Schools and School Districts* (Denver, Co: Education Commission of the States, 2002), <http://ecs.org/clearinghouse/41/52/4152.pdf>.
- ³⁵ Brady, *Failing Schools*.
- ³⁶ Ibid.; New York State Education Department, Office of New York City School Improvement and Community Services, "2002-03 Registration Review Initiative: A Summary of the Registration Review Report Findings" (2003).
- ³⁷ Maryland State Department of Education, "Fact Sheet #5."
- ³⁸ Interview, Maryland State Department of Education official, December 2003.
- ³⁹ Eva Travers, *Characteristics of Schools Under Diverse Providers 2002-2003* (Philadelphia, Pa: Research For Action, 2003).
- ⁴⁰ For an analysis of one EMO's uneven results, see Gerald Bracey, "The Market in Theory Meets the Market in Practice: The Case of Edison Schools," *EPSL: Education Policy Research Unit (EPRU) EPSL-0202-107-EPRU* (2002), <http://www.asu.edu/educ/epsil/EPRU/documents/EPRU%202002-107/EPSL-0202-107-EPRU.htm>.
- ⁴¹ Hess, "Reconstitution."
- ⁴² Jennifer O'Day and Kara Finnigan with David Wakelyn, *External Support to Schools on Probation: Getting a Leg Up?* (Chicago, Il: Consortium on Chicago School Research, July, 2003), <http://www.consortium-chicago.org/publications/p63.html>.
- ⁴³ Ibid. For a description of Philadelphia's early experiences working with external providers, see Blanc's "Supports and Challenges." Also, for studies of Comprehensive School Reform Design implementation, see Joseph Murphy and Amanda Datnow, eds., *Leadership Lessons from Comprehensive School Reforms*. (Thousand Oaks, Ca: Corwin Press, 2002).
- ⁴⁴ For a comprehensive review of the recent research on charter schools, see Katrina Bulkeley and Priscilla Wohlstetter, eds., *Taking Account of Charter Schools: What's Happened and What's Next?* (New York: Teachers College Press, 2003).
- ⁴⁵ The Brown Center, "Charter Schools."
- ⁴⁶ Interview, Research For Action researcher, December, 2003.
- ⁴⁷ Liam Garland, *Navigating Treacherous Waters: A State Takeover Handbook* (Lanham, Md: Scarecrow Education Press, Inc.: 2003).
- ⁴⁸ Todd Ziebarth, *State Takeovers and Reconstitutions* (Denver, Co: Education Commission of the States (ECS), revised 2002).
- ⁴⁹ Garland, *Navigating Treacherous Waters*.
- ⁵⁰ For reflections on the success of North Carolina's Assistance Teams, see Ladd and Zelli, *School-Based Accountability*; and Interview, North Carolina Department of Public Instruction official, December, 2003. For descriptions of California's experiences with School Assistance and Intervention Teams, see Posnick-Goodwin, "Under Close Scrutiny;" and Interviews, School Assistance and Intervention Team Providers, December, 2003.
- ⁵¹ Interview, North Carolina Department of Public Instruction official; Holdzkom, *Low-Performing Schools*; SERVE, *Assisting Low-Performing Schools*.
- ⁵² Interview, North Carolina Department of Public Instruction official.
- ⁵³ Interviews, School Assistance and Intervention Team Providers. For information about California's providers, see California Department of Education, "School Assistance and Intervention Team Providers: Immediate Intervention/Underperforming Schools Program," California Department of Education, <http://www.cde.ca.gov/i/iusp/sait.html> (accessed January 16, 2004).
- ⁵⁴ Ferguson, *The Progress of Education*.
- ⁵⁵ Mintrop, *The Limits of Sanctions*.
- ⁵⁶ Heinrich Mintrop, "State Oversight and the Improvement of Low-Performing Schools in California," expert witness report for *Eliezer Williams et al. v. State of California*, filed at the Superior Court of the State of California, San Francisco, April, 2002. Available at http://128.121.78.24/decentsschools/expert_reports/mintrop_report.pdf.
- ⁵⁷ Laura Goe, "Implementation of California's Immediate Intervention/Underperforming Schools Program: Preliminary Findings" (presentation, annual meeting of the American Educational Research Association,

Seattle, WA, 2001); California Department of Education, "Public School Accountability (1999-2000): Immediate Intervention/Underperforming Schools Program (II/USP): How Low Performing Schools in California are Facing the Challenge of Improving Student Achievement" (2001).

⁵⁸ O'Day and Bitter, *Evaluation Study*.

⁵⁹ Kentucky Department of Education, "Assistance for Schools."

⁶⁰ Jane David, Patricia Kannapel, and G. Williamson McDiarmid, *The Influence of Distinguished Educators on School Improvement: A Study of Kentucky's School Intervention Program* (Lexington, Ky: Partnership for Kentucky Schools, 2000); Holdzkom, *Low-Performing Schools*; Patricia Kannapel and Pamela Coe, "Improving Schools and School Leaders," in *All Children Can Learn: Lessons from the Kentucky Reform Experience*, ed. Roger Pankratz and Joseph Petrosko, (San Francisco: Jossey Bass).

⁶¹ David, Kannapel, and McDiarmid, *The Influence of Distinguished Educators*.

⁶² North Carolina ran a program along similar design features.

⁶³ Deinya Phenix et al., "Evaluation of New York City's Chancellor's District: A Systemic Response to Chronically Low-Performing Schools" (presentation, annual meeting of the American Educational Research Association, Chicago, IL, April 25, 2003); and Interview, New York University's Institute for Education and Social Policy researcher.

⁶⁴ Interview, New York University's Institute for Education and Social Policy researcher.

⁶⁵ Deinya Phenix et al., "Chancellor's District;" and Interview, New York University's Institute for Education and Social Policy researcher.

⁶⁶ Texas Education Agency, "2002 Comprehensive Annual Report."

⁶⁷ Kentucky Department of Education, "Commonwealth Accountability Testing System (CATS): Briefing Packet," 2000.

⁶⁸ Compared to Kentucky, the North Carolina system, with growth expectations pegged to average state growth, yielded a smaller number of identified low-performing schools from its inception. When the state began its ABC tests in the 1996/'97 school year, 123 K-8 schools were identified (7.5% of total). A year later, that number was reduced to only 15 low-performing K-8 schools (0.9%). In subsequent years, the numbers remained low, though they rose again to 44 schools in the 1999-2000 school year, with high schools now being included. But this still constituted no more than about 2% of all schools. (North Carolina Public Schools, "2000-2001: A Report Card for the ABCs of Public Education," 2002)

⁶⁹ Lauren Resnick, Oral communication at the Annual Conference of the American Educational Research Association, San Francisco, 2006.