

**Systemic Strategies to Improve Low-  
Performing Schools—Lessons from  
First-Generation Accountability Systems**

CSE Report 617

Heinrich Mintrop  
CRESST/University of California, Los Angeles

Rosie Papazian  
UCLA

December 2003

Center for the Study of Evaluation  
National Center for Research on Evaluation,  
Standards, and Student Testing  
Graduate School of Education & Information Studies  
University of California, Los Angeles  
Los Angeles, CA 90095-1522  
(310) 206-1532

Project 1.1 Comparative Analyses of Current Assessment and Accountability Systems  
Heinrich Mintrop, Project Director, CRESST/UCLA

Copyright © 2003 The Regents of the University of California

The work reported herein was supported under the Educational Research and Development Centers Program, PR/Award Number R305B60002, as administered by the Institute of Education Sciences, U.S. Department of Education.

The findings and opinions expressed in this report do not reflect the positions or policies of the National Institute on Student Achievement, Curriculum, and Assessment, the Institute of Education Sciences, or the U.S. Department of Education.

**SYSTEMIC STRATEGIES TO IMPROVE LOW-  
PERFORMING SCHOOLS—LESSONS FROM  
FIRST-GENERATION ACCOUNTABILITY SYSTEMS**

**Heinrich Mintrop**

**CRESST/University of California, Los Angeles**

**Rosie Papazian**

**Bay Area Coalition for Equitable Schools**

Failing public schools are a national problem. Highly publicized reports and manifestoes have repeatedly put the spotlight on performance deficits in American schools. In the last 20 years or so, increasing numbers of states and local districts have responded to this problem by creating standards-based accountability systems that are “high stakes,” in the hope that such systems provide incentives for educators to improve their performance. With the passage of the No Child Left Behind (NCLB) Act in 2001, accountability measures have become a cornerstone of the federal government’s approach to schools serving children in poverty, potentially making high-stakes accountability a compelling and pervasive feature in schools all over the United States.

**NCLB and First-Generation Accountability Systems**

According to NCLB, states are to create accountability systems by formulating standards, testing students regularly, defining a baseline and a level of proficiency from 2001 performance levels. Schools are required to attain “adequate yearly progress” (AYP) toward proficiency. AYP can vary from year to year, but all schools need to have reached proficiency for 100% of their students by the school year 2013-2014. Schools that lag behind are subject to an intervention process constructed in three stages: improvement, corrective action, and restructuring. When a school fails to make AYP two years in a row, it enters the improvement stage. Schools in this stage engage in a process of internal school renewal. They write a school improvement plan and implement effective programs, comprehensive school improvement models, and extended services. Districts are required to provide assistance. A school can contract with third-party providers. Parents have the option to enroll their children in another school and upon the school’s failure to

make AYP in the first “improvement” year, parents have the right to enroll their children in tutoring services provided by the district or other organizations. If schools fail to make AYP yet another year, they enter the stage of corrective action during which district intervention intensifies. Among other measures, staff can be removed, curricula mandated, management authority revoked, and instructional time extended. Should a school linger and fail to make AYP yet one more year, major restructuring is to occur via reconstitution, state takeover, conversion into a charter, transfer to a private management company or other, similarly radical measures. Thus, a school that fails to improve for five consecutive years ceases to exist in its original form according to NCLB. Districts encounter a similar staged approach, according to NCLB. When they fail to make district AYP for 2 consecutive years, they enter the improvement stage, which primarily entails programmatic changes. After another 2 years of missing AYP, they are subject to corrective action, which may severely curtail their authority.

As of the summer of 2003, all states have submitted an NCLB implementation plan to the U.S. Department of Education. For many states, the NCLB high-stakes three-stage approach to low-performing schools is novel. But many state governments acted prior to federal legislation. As of the year 2001, 27 states have had school accountability systems that identify low-performing schools; 14 states have stipulated more severe penalties when an underperforming school fails to improve (Boser, 2001). However, only nine states finance student remediation, and seven states provide remedial funding to low-performing schools, according to the ratings of *Quality Counts*, a project of *Education Week* (Quality Counts 2001, 2001).

Some jurisdictions identified quite a substantial number of low-performing schools. For example, in 1997 the city of Chicago alone identified a hundred or so public schools on probation that managed to have fewer than 15% of their students read at the national norm, as measured by the Iowa Test of Basic Skills (Chicago Public Schools, 1997). Between 1995 and 2001, the small state of Maryland identified 100 or so schools statewide. Between 1999 and 2001 alone, the large state of California identified 1,290 persistently low-performing schools that are enrolled in the state’s Immediate Intervention/Underperforming Schools Program. What is more, these schools were not randomly or evenly distributed across the states, but in many instances have been clustered in districts that traditionally serve poor and disadvantaged minority populations. For example, in Maryland, almost all identified schools were located in two districts; in California, 54 of the 1,000 or so

school districts with more than 10 schools had at least a third of their schools eligible for California's Immediate Intervention/Underperforming Schools Program.

The proliferation of high-stakes accountability systems in the 1990s created a new category of low-performing schools prior to NCLB. These schools are presently identified with various labels and they are subject to an array of interventions and sanctions. Although most of these earlier first-generation high-stakes systems echo the structures of NCLB in its basic format, they differ widely in their concrete approaches with repercussions for identified low-performing schools and districts. States implementing NCLB or aligning their already existing accountability system to NCLB may learn from these variations. Although the NCLB accountability regime differs in many respects from the earlier systems, NCLB leaves room for states' discretion, and in many states NCLB implementation is a melding of traditional state structures and new federal demands. Insights from first-generation systems can help avoid less promising design features or suggest likely trajectories for certain system designs.

### **The Research**

This report looks at five state systems, one district system, and one specific program within a state system. We looked at three smaller states (Kentucky, Maryland, North Carolina) and two larger ones (California, Texas). These five state systems constitute the main body of our research. We also looked at New York's Schools Under Registration Review (SURR) program and Chicago's approach to low-performing schools. Data from these programs supplement the findings from the five states in appropriate sections of the report. We selected these systems for four reasons: they are truly first-generation systems in that they have spearheaded high stakes accountability in the U.S. and been in existence for some time; they are fairly elaborated in terms of implemented design features; they have figured prominently in the public discussion on high stakes accountability prior to NCLB; and research material about these systems is available. We examine the systems in the way they were structured prior to NCLB adjustments. Even though some of the systems were redesigned (for example Maryland and Kentucky) over time, second- or third-generation system designers can learn from their experiences.

We asked the following research questions:

- What performance demands did these systems place on schools and/or districts?

- What entrance and exit requirements did low-performing schools programs have in these systems?
- On what scale did they operate?
- What entities were the recipients of accountability?
- What was their design stability over time?
- How did they use sanctions?
- What kind of capacity building did they provide and what management structures did they use for the provision of these services?
- How successful have the programs been in raising test scores and exiting schools in the states' own terms?
- What lessons could be gleaned from the answer to these questions for states that are in the process of designing accountability systems in the wake of NCLB?

Our data are studies, papers, reports, and information from web sites, and we relied on personal communication with officials to fill in gaps. Although we now have reports on the impact of high-stakes testing on schools in several states (Herman, in press; Stecher, Barron, Chun, & Ross, 2000; Stecher, Barron, Kaganoff, & Goodwin, 1998), systematic evaluations of low-performing schools programs are rare. Our more descriptive analysis cannot compensate for this lack. It is generally very difficult to determine the effectiveness of a given program, even more so the effectiveness of a particular element of the program design. Many factors mediate the influence of a particular state policy on school performance, among them district responses and school organizational capacity. It is even more difficult to assess the effectiveness of a program relative to other programs that differ among each other without a common metric that would allow us to compare it in a straightforward way. Given that program design differences structure different performance challenges for schools and different intervention burdens for states, we cannot evaluate the policies with our research. But we attempt to do more than just merely describe design features. In the report, we refrained from burdening the reader with too much descriptive information (see a good descriptive report by Council of Chief State School Officers, 2003). In concentrating on "Lessons Learned," our overview may help systematize and categorize the states' various strategies and their consequences and thus foster a reasoned discussion about low-performing schools

programs and NCLB implementation based on the experience of first-generation accountability systems. Thus we stress lessons learned.

### **System Profiles**

The accountability systems examined in this paper basically employ three mechanisms to achieve their ends. There are mechanisms that identify low performers, motivate them to improve, and leverage resources and support. All the systems examined here address these three functions. But they do this very differently. Some systems put high demands on schools by either testing student achievement with cognitively complex tests or by expecting growth that is boldly set according to an ambitious performance ceiling that all schools are envisioned to achieve within a given period of time. Others take a more moderated approach. They use, for example, basic skills tests that only challenge schools at the lower end of the spectrum, or they set flexible growth targets that are adjusted to the system's current real growth or performance baseline. Some systems only enter schools into the low-performing schools programs that are rock-bottom performers, others identify schools on various absolute performance levels that missed their growth targets. Programs differ on what kind of growth it takes for a school to exit the program and to shed the low-performance label. Moreover, some accountability systems have implemented direct district accountability; others have not. These mechanisms produce low-performing schools programs with different improvement challenges and on different scales (defined here as the number of low-performing schools relative to the total number of schools in the state). Naturally, programs with high performance demands that identify large numbers of schools in the lowest performing category face a higher burden than programs with modest instructional demands that operate on a small scale.

Accountability systems attempt to motivate identified low-performing schools (LPS), and sometimes districts, in three ways: through pressure, meaningful goals, and the provision of resources and support. In some systems, pressures and sanctions are severe, at least on the books; others emphasize capacity building more strongly. Capacity building is organized in various ways. Some states (and districts) provide assistance and oversight to schools that are directly targeted, specifically tailored to the problems of low-performing schools, sustained over time, and directly under the supervision of the state. Others leverage support and assistance in an indirect or unspecified way, for example through regional service centers open to

all schools in need of assistance. States differ as to grants of additional resources to identified schools. Table 1 gives an overview of the programs' prevalent features.

### Lessons Learned

In their own terms, practically every program we examined here claims to be successful and provides data or information that may suggest it is. Most state accountability systems show an increase in test scores statewide over a number of years. Certain fluctuations notwithstanding, almost all of the examined low-

Table 1  
Profiles of Low-Performing School Programs in the U.S.

	Goals	Type of LPS	Scale	Sanctions	Additional resources	Capacity building	Recipients
Kentucky	Cognitively complex; Fixed growth targets; Ceiling for all, deadline extended	All levels, emphasis on lowest performing	Initial High: 20% Low: 7.5%	Used very sparingly; assistance emphasized	Modest	Targeted, specific, sustained, directly supervised (DE/HSE program)	Mainly schools
N. Carolina	Cognitively complex; Flexible growth targets	Lowest performers	Initial High: 7.5% Low: 1%	Very sparingly	None	Targeted, specific, sustained, directly supervised (Intervention teams)	Schools and districts
Maryland	Cognitively complex; No fixed AYP; Ceiling for all, deadline extended	Lowest performers; State discretion in selecting schools	About 7.5% over five years	Very sparingly	Modest	Left to district (indirect support); State monitors	Schools (Districts indirectly)
Texas	Cognitively basic; Flexible minimum standard	Below threshold of required minimum competency	Ranging between 2% to 4%	Very sparingly	None	Indirect and unspecific (Left to district and regional support centers)	Schools and districts
California	Changing tests; Fixed growth targets; Ceiling for all	Below state average, missed growth target	About 20 %	Too new to know	Substantial grants to schools	Indirect (Left to district and third-party consultants)	Schools (District accountability ambiguous)

performing schools programs show a reduction of the number of identified schools over time. These increases and reductions respectively diminish, in some cases to quite an extent, if one excludes from the comparison the first few years of a system's existence when higher numbers of identified schools can be expected and facile adaptation to a new test can yield large increases in performance (Linn, 2000). We have but a few studies that have either validated state test score increases or disputed such increases (Amrein & Berliner, 2002; Carnoy, & Loeb, 2003; Grissmer, Flanagan, Kawata, & Williamson, 2000; Haney, 2000; Klein, Hamilton, McCaffrey, & Stecher, 2000). But there are not enough specific research or evaluation reports available that warrant definitive claims as to the effectiveness of particular strategies or designs used in the examined low-performing schools programs. An exception is the evaluation report on the California low-performing schools program recently published by O'Day and Bitter (2003). Nevertheless, a number of lessons, cautionary in nature, can be learned from the cross-system analysis. These lessons are summarized as follows:

- Less ambitious systems are more stable.
- Sanctions are not the fallback solution.
- Districts need to be centrally involved.
- Even small intervention burdens require a developed capacity building structure.
- Capacity building is key, can take many forms, but should be clearly focused.
- The need for effective instructional programs ought to be balanced with work on professional norms and teacher commitment.

### **Less Ambitious Systems Are More Stable**

Some of the early accountability pioneers constructed their systems with a bold equity vision in mind. Not only were lower performing students to catch up with higher performing ones in a relatively short period of time, they should also catch up as able and complex thinkers with the type of knowledge that would make them successful in the information economy. But this bold equity vision required revision. Kentucky and Maryland, the two accountability systems touted throughout the 1990s for their cognitively complex tests and their ambitious ceilings to be reached by all schools within a relatively short time frame, experienced major

overhaul. The performance-based features of their main assessments were reduced and the new tests have become more traditional, though they are still more cognitively complex (we only looked at the Kentucky tests) relative to some of the other accountability systems examined here. At the same time, the time line for reaching the performance ceiling, originally set as an act of faith rather than based on actual growth targets achieved by schools, had to be revised and extended. It is unfortunate that in these two pioneering systems cognitive complexity, desirable for its educative effects on student learning, was coupled with unrealistic growth expectations. More flexibility in the calculation of growth targets could have perhaps kept up momentum for the states' original educational goals.

Although the California system is still rather novel, it may steer in a similar direction as the early pioneers. Its fixed ceiling, inflexible growth targets, and inclusion of all below-average schools as potentially eligible for the low-performance designation have fast produced a large proportion of schools identified as low-performing (20%), similar to Kentucky in the initial years. But while Kentucky reduced the complexity of the tests over the years, California is in the process of increasing it by introducing standards tests and de-emphasizing off-the-shelf norm-referenced tests in the calculation of the school performance index. The design of the California system is ambitious in setting goals that may lessen achievement gaps. But if the fate of Kentucky's and Maryland's early systems are any indication, adjustment in the state's growth expectations might be necessary or alternatively the state may come to the realization that measures beyond outcome-based accountability systems are needed to address conditions of inequality. Retrenchment has already happened. Given the present fiscal situation, the state has (at least temporarily) curtailed the broad scope of its program and now concentrates on the schools performing in the lowest decile.

By contrast, systems such as the one in Texas (or similarly in Chicago) fared better, but limited their goals. These systems started out with performance demands that were modest and seemingly within the "zone of proximal development" of the lower performing half of the system. For example, in Texas, schools in which fewer than 50% of students (plus identified sub-groups) passed the state tests (TAAS) originally qualified for the low-performance designation. Recently this threshold was raised to 55% (with the exception of social studies). In 2001, 82% passed all TAAS sections statewide. Thus, the state identifies its lowest performing schools, but exit out of the low-performance status seems within reach. Expected minimum

competency thresholds are gradually raised as schools respond to pressures. But accountability pressures are exerted around a basic skills test. This approach has been controversial among researchers. Some have hailed it as progress in educational equity (Sklar, Scheurich, & Johnson, 2000); others have severely criticized it for eroding excellence in teaching and fostering educationally questionable simplification of learning (McNeil, 2000).

Accountability according to NCLB is positioned in between all these systems. It sets a firm ceiling that needs to be reached within 10 years, but its goal is proficiency, and the definition of proficiency is left up to the state governments. States are faced with a dilemma. In light of earlier constructions of accountability systems, it seems advisable for states to pursue modest goals, but with persistence, since NCLB does not provide flexibility in its performance ceiling. On the other hand, states ought not refrain from rigor in their testing systems and pursue cognitively complex assessments. But the experience of the early Kentucky and Maryland systems seems to suggest that it may be difficult to accomplish the swift closing of the achievement gap on a high level of cognitive complexity within the short time frame foreseen by NCLB. Bold equity goals of the sort that inspired the early pioneers may require policies and efforts that go beyond the means of primarily outcome-based accountability systems.

### **Sanctions Are Not the Fallback Solution**

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. Schools could encounter relatively mild public stigma due to the negative performance label imposed on them, more intense scrutiny from review and evaluation teams, more administrative requirements, such as the writing of a school improvement plan, or more severe sanctions. Practically all of the sanctions suggested by NCLB have been on the books or been tried by the systems examined here, though each system's mix may differ from NCLB. In California, principals and teachers are threatened to be reassigned. Schools can be taken over by the state. They can be reorganized, closed, or assigned to the management of another educational or non-profit institution. Parents can select a different public school or apply for charter school status. State take-over is the most severe sanction in the Maryland system. Public hearings, appointment of a special on-site monitor or master, and eventual school closure are envisaged by the Texas regulations as sanctions. Assignment of an instructional officer, external partner,

removal of the principal, and school reconstitution (i.e., staff reassignment and reorganization) figure prominently in the Chicago system. Redesign and closure are also primary sanctions in the New York SURR program (Ascher, Fruchter, & Ikeda, 1999). Kentucky and North Carolina add more personally severe penalties to this list. In addition to the assignment of intervention teams or managers, teachers in low-performing schools can be required to take a general knowledge competency test in North Carolina, and in Kentucky they can be evaluated with the possibility of transfer, demotion, or dismissal. These personally severe sanctions are not part of the NCLB regime.

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.” Only the lowest-performing schools (30 out of the 90 schools “in need of assistance” in 2001) were required to accept assistance. The other 60 had the option to participate. 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. Actual imposition of final sanctions is a negligible feature in Kentucky.

In Texas, more severe sanctions were used very sparingly. 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. The state has reconstituted three schools, all located in one urban district (Ferguson, 2000). The state primarily relies on the threat of bad publicity to motivate districts and schools to improve performance (Izumi & Evers, 2002). Likewise Maryland, after 5 years of high stakes accountability, finally took over three schools and assigned them to private management organizations.

In New York and Chicago, more severe sanctions played a greater role. Within New York’s SURR program, affecting primarily New York City, 35 schools have been closed since the inception of the program. In Chicago, seven schools were reconstituted in the 1997/’98 school year, but this has not been repeated. Moreover, school principals are now receiving training and support from an “area instructional officer” making the original “probation manager” superfluous.

When the present California accountability system was designed, the turn from pressure to support that earlier accountability system was evident. The California

program already began with voluntary participation of qualifying schools, though in actuality most schools were “volunteered” by their districts (O’Day & Bitter, 2003). Schools selected into the program accept increased scrutiny and accountability from the state in return for funds usable for capacity building at the site. Although large proportions of eligible schools that chose not to apply were left out, those that did enroll pinned their hopes for improvement on additional support. The threat of further sanctions was a mere background feature of the program, according to O’Day & Bitter as well as data collection by the author. Likewise, the state itself put only minimal structures in place that could administer additional pressures in cases where the grant making strategy failed. When fewer schools than envisioned met growth targets, the state refrained from building up pressure. It readjusted growth expectations and added additional intervention layers preceding more severe sanctions. To date, out of the first cohort of 430 schools accepted into the program, the state identified 24 schools that require this additional intermediate intervention.

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 (Brady, 2003). This is plausible, but other research (CPRE, 2001; Malen, Croninger, Muncey, & Redmond Jones, 2002; Mintrop, 2003a; O’Day & Bitter, 2003) suggest that, political costs notwithstanding, the pressure strategy is a double-edged sword and not as promising as perhaps originally perceived. This may be so for a number of reasons:

- The results of more severe sanctions and the implementation of major school redesigns have shown to be inconclusive. For example, in California previously locally reconstituted schools in the city of San Francisco showed up again on the state’s low-performing schools list (author’s analysis). In Maryland, some local reconstitutions actually exacerbated schools’ capacity problems, reduced schools’ social stability, and did not lead to the hoped for improvements (Malen et al., 2002). Results from Chicago’s reconstitutions were inconclusive as well (Hess, 2003). In New York, where sanctions were used more vigorously, more than a tenth of the schools were closed, but only about half of the 243 SURR schools have exited the program successfully so far.
- Educating children is a highly complex task, but high-stakes accountability systems usually privilege very few performance indicators, often one central test for instructional performance. Even though these systems are to focus teachers on instruction, attaching too much pressure to a single indicator may have undesirable consequences. Teachers teach to the test, producing test training effects, rather than substantive learning effects (Linn, 2000). Moreover, forcing teachers to unduly narrow the scope of their work creates serious acceptability problems for the state assessments, which

are then perceived as invalid indicators of a teacher's work. As a result, the educational meaningfulness of accountability systems among pressured teachers may be low and accountability goals lose motivating force (Mintrop, 2003a; Mintrop, 2003b).

- Heightened pressure exacerbates already severe teacher commitment problems in many low-performing schools. Many low-performing schools are not attractive workplaces, and under current labor market conditions, schools in many jurisdictions with high concentrations of low-performing schools are staffed with large numbers of new, often insufficiently trained teachers with low commitment to stay. Likewise, principal turnover is high as well. Principals under pressure of accountability often act as conduits of pressure, making for unsupportive working relationships between teachers and administration. Thus, too much pressure may lead to dissatisfaction, exit, or additional organizational fragmentation (Ingersoll, 2001; Malen et al., 2002; Mintrop, 2003a). Potential staff replacements are not necessarily of higher quality than the original teaching staff.
- Identifying low-performing schools has put the spotlight on glaring capacity deficits in these schools that a motivation strategy alone cannot remedy (Finnigan & O'Day, 2003). This in turn brings issues of fairness and attribution to the fore (Mintrop, 2003a). For example, in many accountability systems, districts are either named or not targeted as direct recipients of accountability measures, but district context and district (as well as state) policies are often times critical of the school's learning conditions and ultimate success (CPRE, 2001). When schools and teachers feel forced to assume accountability for critical conditions of student performance over which they lack authority and control, they may reject accountability altogether, rather than assume responsibility for their contribution to the problem. In this case, accountability becomes counterproductive and demotivating.

In summary, all accountability systems use mild pressure as a means to motivate educators to improve performance, but the use of more severe sanctions is de-emphasized. In their majority, the examined states have either hardly ever used or turned away from high pressure as a main lever to motivate teachers. Under current conditions, high pressure and severe sanctions may be often counterproductive. The identification of low-performing schools has revealed a crisis of capacity that state support strategies need to address.

A central feature, if not *the* central mechanism, for school improvement under NCLB is the three-staged process of increasing high stakes through improvement, corrective action, and restructuring. Under NCLB, schools may face severe sanctions in a rather short time, and voluntary participation is excluded as an option. If experiences of the first-generation accountability systems are any indication, states

are advised not to rely on the power of pressures and sanctions to get the job of school improvement done. Rather, states need to construct low-performing schools programs that place heavy emphasis on support and intervention, bolster commitment of teachers to low-performing schools, and strongly motivate educators. Such accountability systems set goals that are deemed realistic, use assessments that are educationally meaningful (i.e., deemed valid and fair), facilitate school evaluations that allow schools to see their contribution to the performance problem, offer suggestions on how schools can improve, and identify those barriers of performance that district and state policies are called to remedy.

### **Districts Need to Be Centrally Involved**

A number of first generation accountability systems were designed with the idea of freezing out middle management (i.e., districts) by holding schools directly accountable to central state goals. This was done most radically, not in the U.S., but in England where local educational authorities lost most of their authority and schools received control over 90% of their budgets. In the U.S., districts for the most part retained their power. Some accountability systems aimed to influence district action indirectly by targeting individual schools.

The Maryland system is a good example. Maryland has few regulatory mechanisms to hold districts directly accountable within the purview of the low-performing schools program. (Districts highly impacted by low-performing schools write a master improvement plan.) Schools are the units that are to absorb the blow of potential sanctions, but districts are expected to provide the necessary conditions for these schools to be successful. In the Maryland case, the state tried to influence districts indirectly. The state identified schools almost exclusively in two districts. In Baltimore City, almost half of the schools were designated. This provided the stage for the state to press on more far-reaching changes in district governance and administration. High exit requirements in conjunction with low numbers of exiting schools kept up these pressures on the local districts (Cibulka & Lindle, 2002). In Kentucky, the state program identifies schools, not districts, and provides all services that come with this designation centrally and directly to schools. Originally, the accountability system was designed with the idea of overriding detrimental district influences by establishing a direct relationship between schools and the state. But more recently, the state has begun to reach districts more directly by providing training and support to district personnel. Even the much younger California accountability system has been fairly ambiguous as to the role of districts,

and state interventions in districts has been largely restricted to cases of fiscal impropriety.

By contrast, Texas and North Carolina have had strong district accountability features. In Texas, districts are directly held accountable and failing districts are put under pressure of sanctions. In 1995, for example, the state identified 34 “academically unacceptable” districts (Texas Education Agency (TEA), 2002). That number diminished to one district in 2001, but surged again to 16 districts in 2002. Researchers (Sklar, Scheurich, & Johnson, 2000) point to the importance of district accountability for the effectiveness of the Texas accountability system. North Carolina as well holds districts (superintendents and boards) directly accountable for having at least half of their schools performing with satisfaction. Sanctions are imposed in case of persistent underperformance.

NCLB spells out mechanisms for district accountability, and all states are therefore required to develop intervention and corrective action measures for their districts.

The first generation systems demonstrate a certain equivocation when it comes to districts. But to the extent that the *high stakes* features of pressures and sanctions have shown to be limited and capacity building strategies instrumental, the role of the local district in accountability systems becomes germane. A thorough evaluation of the California low-performing schools program conducted by the American Institutes of Research (O’Day & Bitter, 2003) found that districts had a strong influence on low-performing schools’ improvement trajectories. In fact, when “schools’ district” was entered as an explanatory factor for increasing test scores, the effect of schools’ enrollment in the low-performing schools program vanished in the authors’ statistical models. It behooves states to define carefully and without ambiguity districts’ role as accountability recipients. Particularly in larger states, precious state oversight and intervention capacities might be better spent on districts rather than large numbers of more distant individual schools. State sanctions most often relate to the organization, management, and governance of individual schools. Local districts, on the other hand, can intervene far more deeply into the instructional core of a school, and its policies may often be more influential in shaping school success or failure than those of states

## **Even a Small Intervention Burden Requires a Developed Capacity Building Structure**

Surprisingly, apart from a program's initial stages when the load of identified schools can be up to a fourth of all schools (e.g., Kentucky, California), first generation accountability systems have kept their intervention burden fairly modest. A load of 7 to 8% can be considered high, as in the case of Kentucky. But the experience of this state and others show that even this low number of schools requires an elaborate system of support and intervention.

In the 1996 to '98 biennium, the second biennium of the original highly complex accountability system, Kentucky entered 250 schools into the low-performing schools program (Cibulka & Lindle, 2002). With roughly 1,200 schools in the state, this constituted more than 20% of all schools. But these schools were not necessarily academically failing. They had growth deficiencies, some on high absolute performance levels. Most of the 250 schools did not continue in the status. But their exit coincided with a redesign of the system (Kentucky Department of Education (KDE), 2000a; 2000b), making judgements of effectiveness difficult. In the 2002 accountability cycle, the state has identified 90 schools as low performing or about 7.5% of the total (KDE, 2002).

Compared to Kentucky, the North Carolina system, with growth expectations pegged to average state growth, has 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, 2002). Thus, the North Carolina situation is characterized by a relatively light load of low-performing schools that has consistently been held low. Nevertheless, the state's support structure is elaborate and intensive.

Although more numerous in absolute terms, the Texas low-performing schools program is relatively small scale as well. In 1995, the system identified 267 low-performing schools. The numbers dropped to 59 in 1998, and rose again continuously to 150 in 2002. In 2002, the program identified 150 schools (TEA, 2002), but the thresholds for entrance and exit rose in the meantime. With these numbers,

the program fluctuated in the 2 to 4% range of the total number of schools in the state. Thus in Texas, relatively small proportions of schools have been identified as low performing. At the same time, the state's support and intervention system is less elaborate than that of the states previously discussed, but the demands on schools seem also more modest and exit criteria in easier reach.

Among the examined systems, the toughest challenge ahead was created by the Maryland system. 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 very high. The state department limited the burden of the low-performing schools program by capping the number of schools at around a hundred (about 7% of all schools) although more schools could have qualified according to the state's criteria, but one district is burdened with managing about half of its schools as identified low performers. The state did not develop an elaborate capacity building structure. Very few schools have managed to exit the program as of this writing. (The revamping of the Maryland system occurred simultaneously with the research for this report and is not reflected here.)

California's relatively novel program experienced a surge of identified low-performing schools. Growth expectations and entrance rules for below-average performers were set in such a way that after 3 years the low-performing schools program<sup>1</sup> has enrolled about 20% of all schools that received an Academic Performance Index, or about 1,500 schools. The scale of the program is curtailed by its voluntary feature. Being voluntary, schools and districts decide whether they will apply for additional funds in return for scrutiny and threats of further interventions. In 2001, only 527 or 56% of the 935 eligible schools applied. In 2002, of the 1,266 eligible schools, only 765 or 60% applied to the program, thus about half of the eligible schools decided to by-pass the program each year. The state ended up accepting only 430 schools each year for funding (through its main program). Had all eligible schools been designated, the scale of the program would have been enormous. California's program is still in its nascent state, and so are support and intervention systems, but it appears that out of the first cohort that entered the program only about one fourth of the schools managed to meet the state's growth expectations, qualifying them for outright exit.

---

<sup>1</sup> There are actually two programs, the main one applying to all schools below the 50<sup>th</sup> percentile, the secondary one targeting Decile 1 schools.

First generation systems have found ways to keep their programs surprisingly small either by cutting off identification (Maryland) or readjusting demands (Kentucky). These strategies will no longer be possible under NCLB as long as the federal government itself does not make adjustments. Even states with small intervention burdens (but ambitious goals) have found it necessary to build up an elaborate capacity building structure. States without these structures had difficulties exiting schools if performance demands were set high at the same time. As NCLB requires schools to make swift progress once they are identified by the state's accountability system, states need to prepare for a potentially high improvement and intervention burden for which structures and capacities need to be developed.

### **Capacity Building Is Key, Can Take Many Forms, But Needs to Be Focused**

Research by O'Day and Bitter (2003) found that schools' responses to being in the low performance program varied widely. Data from Chicago suggest a similar pattern. Some schools improved rapidly while others lingered in the program. Initial capacity was a key factor in explaining these results. Elementary schools with higher initial capacity that had higher "peer collaboration, teacher-teacher trust, and collective responsibility for student learning" (O'Day, in press, pp. 2-27) responded more favorably. These findings point to the critical importance of capacity building in low-performing schools. It seems that schools with low organizational capacity benefit only weakly from accountability policies. But it also appears that even relatively intense provision of support, as in the Chicago case, was not sufficient for a sizable number of (elementary) schools to exit the low-performance status.

The low-performing schools programs, examined in this paper, commonly consist of the following elements directly pertaining to the school level:

- **Additional funds:** They are not present in all programs. In some programs the sums are negligible; in others they are substantial.
- **Evaluation/audit:** These can be short, unstructured visits from state department officials or extensive one-week inspections during which the school's operations are examined comprehensively.
- **School improvement plans:** The requirement that low-performing schools write these plans according to state or district templates is a universal feature across all programs. The programs differ in the degree to which these plans are reviewed and validated by an external authoritative body and in the degree to which their implementation is monitored on site.

- On-site personnel: In the most basic version, they are just monitors of the school improvement plan or the general development of the school, the “eyes and ears of the state.” In some programs, they primarily have a helping role. They provide support in analyzing test data, observe lessons and give model lessons, help in selecting instructional programs and instructional strategies, provide staff development, and give management advice. In some programs, they have a more authoritative role as they evaluate teachers and principals, and give reports to governing bodies.

At the state level, the provision of such services needs a management structure that ensures high quality of delivered services. Personnel, be they monitors, evaluators, management consultants, or instructional specialists, needs to be recruited and trained to make sure their work is focused on state goals, yet flexible enough to address school-specific needs. Work quality is insured through regular performance reviews and the opportunity for support teams to meet and learn from their experiences. First-generation accountability systems differ in the degree to which these school level oversight and support services are developed.

**Elaborate state structures.** Of the state programs we surveyed, Kentucky and North Carolina have fairly comprehensive systems in place that provide oversight and support to schools under direct supervision from the state department. Services are sustained over one school year or longer, and specifically targeted to low-performing schools and state goals. As part of the state’s support for its schools “in need of assistance,” Kentucky provides modest additional school improvement funds. In the 2002-2003 year, \$2 million has been budgeted for the 90 schools. For example, elementary school grants range from \$12,000-\$38,000 per biennium. The money is supposed to be targeted to those students that are in most need of improvement.

Schools in the “need assistance” category are subject to an intense self-study and audit that examines their performance on nine standards and indicators for school improvement in the following areas: curriculum, classroom assessment, instruction, school culture, learning environment, family and community involvement, professional growth development and evaluation, leadership, organizational structure and resources, and planning. The review is conducted by state-sponsored Scholastic Audit Teams, which include a highly skilled educator (HSE), a teacher, a principal or other administrator, a parent, and university-based educator (KDE, 2000a). The audit teams are trained for their task.

The audit teams visit each school for about a week. Once the scholastic audit has been conducted, schools use the results to write their school improvement plans. The lowest category of performers (Level III) receives mandated assistance from an HSE for the entire biennium; the others receive voluntary assistance. School plans are written with the help of the designated HSE and are submitted to the state department for review and approval. A state-certified person other than the HSE also conducts an evaluation of school personnel at all Level III schools. Principals at all three levels are required to participate in staff development to enhance leadership skills.

HSE's have to demonstrate prior ability to bring about high levels of student performance and go through a rigorous hiring and training process. Each HSE receives two weeks of training and follow-up training at quarterly meetings. Mentors from the state department provide assistance in problem solving and support to HSE's. HSE's are expected to serve on-site at least 80% of their work time. Their activities include but are not limited to: staff development, classroom observations of instruction, demonstration lessons, grants writing, tutoring, and creation of model lessons (David, Kannapel, & McDiarmid, 2000; Holdzkom, 2001; Kannapel & Coe, 2000). In addition, a team of HSE's that specializes on organizational management has been formed and can be assigned to more than one school at a time, given the needs of a particular school. In the 2002-2003 school year, there are 52 HSE's working with 30 Level III schools and providing support to others on a voluntary basis.

Quite a bit of research has been focused on the effectiveness and impact of the HSE (or as it was previously called Distinguished Educator) program. The majority of it speaks to its success as a capacity building tool in low-performing schools. According to one study, the DE program had a significant impact on test scores and school culture, (David, Kannapel, & McDiarmid, 2000; Kannapel & Coe 2000). A reported key focus of the work of the HSE was curriculum and instructional alignment to the pedagogically complex state assessments. Teachers reported improved preparation for the state assessments, including improved curriculum and instructional coordination at the school, greater attention to test data, and increased information on how to prepare students to succeed on the state test. Test score data show that schools that participated in the DE/HSE program improved at a higher rate than those that did not (Kannapel & Coe), although it is difficult to isolate the impact of HSE's in the whole school environment. Significant challenges for the

program are sustaining the change once HSE have left school grounds, creating an appropriate match between the HSE and the school, and maintaining a strong pool of HSE's (David et al., 2000).

In North Carolina, no additional funds are allocated to low-performing schools, but these schools receive intensive oversight and support. Low-performing schools are assigned an external assistance team made up of one administrator and three or four teachers with experience at the grade span of the school being served. Each team works with a school for one academic year on a daily basis. The teams' tasks are similar to the ones HSE's in Kentucky carry out. In addition, they report to the local school board or the state department on the school's progress.

Assistance team members participate in a 4-week comprehensive training in topics similar to those in Kentucky. In addition, the assistance teams can participate in 2 extra weeks of training in a program specifically designed to reduce minority achievement gaps and are encouraged to go to conferences regarding specific subject areas or grade-level content. A team of five people working at the state level provides technical assistance to assistance teams. In the 2000-2001 school year, the state employed 80-85 assistance team members and served a total of 52 schools, with 14 schools receiving a mandated assistance team. An inquiry by the state department has revealed that assisting schools in data analysis, modeling good instruction, and aligning the schools' curricula to state curricula and assessments has been instrumental in moving schools on the path to improvement.<sup>2</sup>

**Focus of support.** In the city of Chicago, accountability measures were coupled with extensive provision of support and oversight from the beginning. Schools, primarily principals, receive the services of various management consultants dispatched directly by the district as well as the contracted services of external partners that schools select themselves. The district pays for these services, but after the third year schools are expected to pay for the external partner out of their own budgets. The external partners have various programmatic foci, but are all concerned with curriculum and instruction, staff development, and other aspects of school life that affect the staff as a whole. The district offers schools a core course of study to follow, but schools are not required to follow it. Thus, in instructional matters, Chicago leaves discretion to schools and allows for variation in intervention strategies. The district evaluates the performance of external partners regularly, and

---

<sup>2</sup> ([http://www.ncpublicschools.org/school\\_improvement/assistance\\_index.html](http://www.ncpublicschools.org/school_improvement/assistance_index.html))

partners that do not help raise students' test scores above district average are not retained. But a systematic exchange of information among partners on what works is not sponsored, leaving partners free to promote their own products. This, according to research by Finnigan & O'Day (2003) has supported a tendency of rather unfocused provision of support in the area of curriculum and instruction that is not concentrated enough on district concerns, for example in the area of literacy. This research suggests that focused on-site support tied to classrooms and concrete accountability goals may be more forceful in improving schools than staff development dispensed by external consultants with varying programmatic foci.

**Leveraging support indirectly.** Texas, California, and Maryland leverage support for their low-performing schools in different ways. These states expect local districts to step up and provide the necessary interventions once low-performing schools have been identified under their jurisdiction. Only Texas, however, has strong mechanisms built into its accountability system that identifies low-performing districts directly and threatens them with further sanctions.

Texas has a decentralized form of governing schools. Most decisions are made on site, and as a result, the state does not take a very strong leadership role in providing support to ailing schools (Ferguson, 2000) and has only limited capacity to do so. However, the state requires low-performing schools and districts to compile a school improvement plan. It sends peer review teams to schools and districts that visit a school or district for varying lengths of time depending on size of school or district. These peer review teams are made up of state department staff and evaluators that receive training with the help of a CD. Participants need to have special qualifications and must have local superintendent approval to participate. Local districts pay for travel and release time for participants. In addition the state has organized educational support centers that offer their services to low-performing schools and districts, but not exclusively so. Other schools in need of support can contact these centers as well. Texas does not furnish additional monetary grants to low-performing schools. Only a small number of schools are under more direct supervision. As was mentioned above, only seven schools are currently visited by monitors and two schools supervised by so called masters. Thus, perhaps with the exception of the monitor and master feature, the support the state leverages for low-performing schools is more indirect.

As was pointed out, California, not unlike some other systems, started out identifying a large number of schools very swiftly.<sup>3</sup> Identified schools have 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. The state compiled this list based on written applications received from these external vendors or agencies. Training in evaluation was not provided. The state, however, did require vendors to reapply to the list showing evidence of success. The external evaluators negotiate with schools the extent of their fees and services. The state provides schools with a \$50,000 planning grant that can be used to pay the external evaluator, and then another \$ 200 per student over 2 years that is to pay for capacity building measures chosen at the school's discretion. During these 2 years, the school is expected to have met its growth targets. To receive this money, schools are 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 keeps a low profile. It relies 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 facilitating quality assurance of the support system is only weakly developed. Early reports show that schools' initial responses to the program vary widely and depend on the varying quality of external evaluators (Goe, 2001; Just, Boese, Burkhardt, Carstens, & Gaffney, 2001). A systematic evaluation of the program (O'Day & Bitter, 2003) matching schools enrolled in the low-performing schools program with eligible schools that did not enroll showed no lasting effect on test scores for the enrolled schools. Increased accountability pressures in conjunction with modest grants did not move these schools on a more successful improvement trajectory than low-performing schools that did not receive this treatment. It is not clear, however, why this is so. Qualitative data suggest that the schools lacked sustained quality support and intervention.

At this point, the state has made first attempts to put structures in place that could handle a potentially burgeoning intervention burden. For the first cohort, the

---

<sup>3</sup> We only look at the larger II/USP program, and exclude the later HPSG program that specifically targets all rock-bottom performers (decile 1) independently of meeting the system's growth targets.

state, by relaxing growth expectations, chose to concentrate on 24 out of 430 of identified schools. If the state follows the original pattern, such interventions will be covered through the services of third-party vendors with relatively loose oversight or through spot checks by state department personnel akin to compliance reviews. Ultimately, the state will be in need of a more elaborate structure for which it will be forced to build up state capacity that does not presently exist.

**Goals, capacity building strategies, and program scale.** High-quality support and oversight need to be an integral part of a low-performing schools program to complement the limited impact of pressure and sanctions. First-generation accountability systems show that this is even more so the case when performance demands are complex. Some programs handle a fairly modest load of cases, stress support over sanctions, supervise this support centrally, and manage 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 managerial requirements and performance goals. The low-performing schools program operates in the context of an accountability system with complex performance demands and a high level of guidance by way of a state core curriculum. But at the same time, the schools' instructional program is not standardized and on-site support providers adapt their intervention to individual school needs, though curriculum and instructional alignment are key points of intervention. We saw such patterns most clearly in the programs from the small states of Kentucky and North Carolina.

A program design that places demands for high instructional complexity on schools, identifies rock-bottom performers with probably very low organizational capacity to begin with, establishes a high exit threshold, and leaves it up to (overburdened) districts to provide necessary capacity building will run into trouble even when it keeps the statewide load of identified schools fairly small. We saw tendencies of such a pattern in the state of Maryland where very few schools indeed have been exiting the system successfully.

A program design that identifies large numbers of below-average schools on the basis of a set performance ceiling and fixed growth expectations, stresses grant making over accountability, leaves it up to districts to provide capacity building, but does not make these districts direct recipients of accountability measures, relies on a network of external consultants for evaluation and intervention, but has a very weak management structure at the state level in place that could assure quality of services, may run into trouble down the line. This could be particularly the case once the

novelty effect of the system wears off. We saw tendencies of such a pattern in the early implementation of the California accountability system. Indeed only about a fourth of the first cohort of identified schools met the state's originally stated expectations, and overall the program showed no effect on test scores over 3 years. If past accounts of educational reform are an indication, provision of money, even generous grants, without a clear focus on goals and strategies to achieve them will not be very effective. By contrast, Texas refrains from grant making, compels districts into action through strong district accountability and augments with generic regional services, but this rather austere model of capacity building functions in the context of modest and flexible performance demands.

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. This often requires building up state capacity as well. Compared to some of the first-generation accountability systems, the heavy emphasis of NCLB on intervention in the first 3 years of a school's identification highlights the importance of effective intervention structures, especially when coupled with ambitious performance goals.

### **The Need for Effective Instructional Programs Ought to Be Balanced With Work on Professional Norms and Commitment**

Some districts burdened with large numbers of state-identified low-performing schools chose a path of more centralized control and concentrated resources and energy on a few key instructional and organizational strategies to improve schools. These strategies are particularly popular for elementary schools where the learning goals of basic literacy and numeracy are fairly clear; tested programs are available; and standardization is more easily achieved and accepted. New York's Chancellor's District and a similar district in Baltimore are examples. Here, some of the harder cases are taken out of the regular district governance structure. Placed under special supervision, schools' programmatic and organizational choices are constrained, principal and staff behavior is regularly monitored, and student learning is regularly assessed. For example, in the Chancellor's District, schools teach very structured reading and math programs in 90-minute blocks, staff development is specifically geared to the schools' programs, principals are trained and their instructional leadership is supervised, class sizes are reduced, and instructional time is extended. The chancellor's district has succeeded in moving its original elementary schools out of SURR, the program for the lowest performing schools in New York, but it was

less successful for middle schools (Snipes, 2002). In other cities with large numbers of identified low-performing schools, a similar programmatic standardization was attempted. Los Angeles, for example, made a particular reading program mandatory for all elementary schools and backed this up with a system of on-site literacy coaches whose job is both to provide help and enforce compliance with the new policy.

Mintrop, one of the authors of this paper, and his associates studied a number of cases in Maryland (Mintrop, 2003a, 2003b; Mintrop & Associates, 2001). They found that in schools that moved as a result of probation, principal leadership and faculty collegiality and cohesion as well as trust in the skills of colleagues were stronger. By contrast, career anxiety due to pressure was not higher, nor was there a greater sense of fairness about the system and realism regarding the accountability goals. If anything, teachers in the moving schools were more skeptical about the system's validity than in the stuck schools. At the same time, more motivated teachers (and more motivated faculty in the aggregate) were not necessarily more committed to staying at their negatively labeled school. The moving schools, according to case study data, moved not necessarily because accountability or probation was more motivating to them, but because they had authoritative principals who were supportive and controlling at the same time. These principals, curtailing discussion and demanding action, instituted a new regime of classroom monitoring and observation while *highly qualified* on-site instructional specialists, backed up by the principal, offered instruction-related help. These instructional specialists were hired with additional grants furnished by the district to its state-identified low-performing schools. Although teachers complied with their principals' initiatives and increased effort, particularly in light of the school's probation predicament, they resented additional pressure and control. Disposition to leave the school, even among teacher leaders, was high (Mintrop, 2003a; 2003b). By contrast, "creation/renewal of teachers' commitment to the school" is one of the most salient issues a school needs to address, according to an English report that summarized insights from inspection reports on 900 schools "under special measures" (i.e. identified low-performing schools on probation; Gray, 2000).

Ideally, when a school becomes identified as low performing by the state or district, the school would use performance information to recognize its shortcomings, adjust the school's professional norms and expectations to high external expectations, and avail itself of more effective practices that address

diagnosed performance problems. But it seems that probation pressures tend to trigger a dynamic of managerial control in schools that may preempt the kinds of communication processes needed for teachers to adopt accountability goals as meaningful goals for themselves and commit to these goals. Programmatic standardization and increased supervision, as may occur in the special districts, may facilitate the implementation of programs deemed effective by district administrations, but may further constrain the space for the development of what Abelmann, Elmore, Even, Kenyon, and Marshall (1999) call “internal accountability,” the process from which professionals derive commitment to their work.

We see, in sum, that district provision of support and oversight for low-performing schools can intervene deeply into the core of a school’s operation, and in the examples described here, have done so. Two dynamics seem to be possible here, and it is not clear at present how they are playing out at the schools. In one dynamic, the *low-performing schools* label primarily serves as a tool to establish compliance among educators for programmatic changes selected by district administrations and implemented by districts through a strategy of managerial control and targeted capacity building (e.g., district literacy coaches). In the other dynamic, the label serves as a motivator for the school to develop new professional norms, and capacity building strategies at the school bolster the overall organizational capacity of the school with a focus on instruction. While the former may shorten program implementation time, the latter may be more useful in attracting and committing high-quality, more professionally minded teachers to low-performing schools, and in sustaining improvements over time.

Implementation of effective programs is desirable and especially necessary when schools are staffed with many insufficiently qualified teachers. Such implementation, however, raises the specter of compliance, managerial control, and programmatic standardization as the main levers of school improvement in low-performing schools. Implementation ought not come at the expense of developing professional norms of high expectations and conduct. Such norms are not only necessary for teachers to collectively assume responsibility for student learning, but also important in fostering and maintaining teacher commitment to stay, a precarious commodity in many low-performing schools. If organizational capacity, understood as the capacity of individuals to interact with and rely on each other, is a key ingredient for schools to respond positively to performance challenges, then

work on internal organizational norms may be a good way to improve on that front. High external expectations can only become normative in the internal context of schools if teachers have the opportunity to explore the meaning of accountability goals for themselves and if they can identify their own and their school's contribution to the school's problems. This also means that school evaluations provide the tools to carefully distinguish between district, state, school, and family responsibilities.

With respect to instructional program development, qualified instructional specialists lodged in a given school seem to have had promising effects on improvement. By contrast, external consultants' efforts may have a tendency to dissipate. Instructional specialists, but also management consultants, or change agents who cover both areas, need training specific to the conditions of the low-performing school.

Motivation and intervention strategies, together, need to create a balanced effect on instructional programs, educators' professional norms of performance, and commitment to stay in the low-performing school. Pressures on principals or staff to comply with external accountability regimes or programmatic decisions should not occur to the detriment of schools' interaction in processing new demands and teachers' willingness to remain at the negatively identified school. Implementation of effective instructional programs should be complemented with organizational development and teacher commitment strategies that are apt to stabilize the low-performing school.

## References

- Abelmann, C., Elmore, R., Even, J., Kenyon, S., & Marshall, J. (1999). *When accountability knocks, will anyone answer?* (CPRE Research Report Series RR-42). Philadelphia, PA: Consortium for Policy Research in Education.
- Amrein, A., & Berliner, D. (2002). High stakes testing, uncertainty, and student learning. *Education Policy Analysis Archives*, 10(18). Available from <http://epaa.asu.edu/epaa/v10n18/>
- Ascher, C., Fruchter, N., & Ikeda K. (1999). *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.
- Boser, U. (2001, January). Pressure without support. *Education Week*, 20, pp. 68-71.
- Brady, R. (2003). *Can failing schools be fixed?* Washington, DC: Thomas Fordham Foundation.
- Carnoy, M., & Loeb, S. (2003). Does external accountability affect student outcomes? A cross-state analysis. *Educational Evaluation and Policy Analysis*, 24(4), 305-331
- Chicago Public Schools. (1997). *School probation information packet*. Chicago: Office of Accountability.
- Council of Chief State School Officers (CCSSO). (2003). *State support to low-performing schools*. Washington, DC: Author.
- Cibulka, J., & Lindle, J. (2002). *The politics of accountability for school improvement in Kentucky and Maryland*. Tech. Rep. Washington, DC: Office of Educational Research and Improvement, U.S. Department of Education.
- CPRE. (2001). *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.
- David, J., Kannapel, P., & McDiarmid, W. (2000). *The influence of Distinguished Educators on school improvement: A study of Kentucky's School Intervention Program*. Frankfort: Partnership for Kentucky Schools.
- Ferguson, C. (2000). *The progress of education in Texas*. Austin: Southwest Educational Development Laboratory.
- Finnigan, K., & O'Day, J. (2003). *External support to schools on probation: Getting a leg up?* Research Report. Philadelphia: Consortium for Policy Research in Education.

- Goe, L. (2001, Spring). *Implementation of California's immediate intervention/underperforming schools program: Preliminary findings*. Paper presented at the Annual Conference of the American Educational Research Association, Seattle, WA.
- Grissmer, D., Flanagan, A., Kawata, J., & Williamson, S. (2000). *Improving student achievement: What state NAEP test scores tell us (MR-924-EDU)*. Santa Monica, CA: RAND.
- Gray, J. (2000). *Causing concern but improving. A review of schools' experiences*. Research Brief No. 188. London, GB: Department for Education and Employment.
- Haney, W. (2000). The myth of the Texas miracle in education. *Education Policy Analysis Archives*, 8(41). Available from <http://epaa.asu.edu/epaa/v8n41/>
- Herman, J. (in press). Instructional effects in elementary schools. In S. Fuhrman & R. Elmore (Eds.), *Redesigning accountability systems*. New York: Teachers College Press.
- Hess, G. A. (2003). Reconstitution three years later. *Education and Urban Society*, 35(3), 300-327.
- Holdzkom, D. (2001). *Low-performing schools: So you've identified them—now what?* Charleston, West Virginia: AEL.
- Ingersoll, R. (2001). Teacher turnover and teacher shortages: An organizational analysis. *American Educational Research Journal*, 38(3), 499-534.
- Izumi, L., & Evers, W. (2002). State accountability systems. In W. Evers & H. Walberg (Eds.), *School accountability: An assessment by the Koret Task Force on K-12 education* (p. 198): Hoover Institute Press Publication; No. 512.
- Just, A., Boese, L., Burkhardt, R., Carstens, M., & Gaffney, T. (2001). *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*. Sacramento, CA: California Department of Education, Division of Policy and Evaluation.
- Kannapel, P., & Coe, P. (2000). Improving schools and school leaders. In R. P. Pankratz & Joseph (Eds.), *All children can learn: Lessons from the Kentucky reform experience* (pp. 159-176). San Francisco: Jossey-Bass.
- Kentucky Department of Education. (2000a). *Assistance for schools: Guidelines for scholastic audit*. Frankfort, KY: Kentucky Department of Education, Attorney General Office. Available from <http://www.lrc.state.ky.us/kar/703/005/120.htm>.
- Kentucky Department of Education. (2000b). *Commonwealth accountability testing system: Interim accountability cycle briefing packet (1996-1997 to 1999-2000)*. Frankfort, KY: Kentucky Department of Education.

- Kentucky Department of Education. (2002). *Commonwealth accountability testing system (CATS): Briefing packet*. Frankfort, KY: Kentucky Department of Education.
- Klein, S., Hamilton, L., McCaffrey, D. F., & Stecher, B. M. (2000). *What do test scores in Texas tell us?* Washington, DC: RAND.
- Linn, R. (2000). Assessment and accountability. *Educational Researcher*, 29(2), 4-16.
- Malen, B., Croninger, R., Muncey, D., & Redmond-Jones, D. (2002). Reconstituting schools: "Testing" the "theory of action." *Education Evaluation and Policy Analysis*, 24(2), 113-132.
- McNeil, L. (2000). *Contradictions of school reform: Educational costs of standardized testing*. New York: Routledge.
- Mintrop, H. (2003a). The limits of sanctions in low-performing schools. *Education Policy Analysis Archives*, 11(3).
- Mintrop, H. (2003b). *Schools on probation: How accountability works (and doesn't work)*. New York: Teachers College Press.
- Mintrop, H., & Associates. (2001). *Schools on probation: Pressure, meaning, capacity and the improvement of schools*. College Park, MD: Center for Educational Policy and Leadership, University of Maryland.
- North Carolina Public Schools. (2002). *2000-2001: A report card for the ABCs of public education*. North Carolina Public Schools. Available from [http://www.ncpublicschools.org/abc\\_results/results\\_01/numberreceiving.html](http://www.ncpublicschools.org/abc_results/results_01/numberreceiving.html)
- O' Day, J. (in press). Complexity, accountability, and school improvement. In S. Fuhrman & R. Elmore (Eds.), *Redesigning accountability systems*. New York: Teachers College Press.
- O' Day, J., & Bitter, C. (2003). *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.
- Quality Counts 2001. (2001). Quality Counts 2001: A Better Balance. *Education Week*, 20(17), 8-193.
- Sklar, L., Scheurich, J., & Johnson, F. (2000). *Equity-driven, achievement-focused school districts: A report on systemic school success in four Texas school districts serving diverse student populations*. Austin, TX: Charles A. Dana Center.
- Snipes, J., Doolittle, F., & Herlihy, C. (2002). *Foundations for success: Case studies of how urban school systems improve student achievement*. Council of Great City Schools.

Stecher, B. M., Barron, S. L., Chun, T., & Ross, K. (2000). *The effects of the Washington state education reform on schools and classroom* (CSE Tech. Rep. No. 525). Los Angeles: University of California, National Center for Research on Evaluation, Standards, and Student Testing (CRESST).

Stecher, B. M., Barron, S. L., Kaganoff, T., & Goodwin, J. (1998). *The effects of standards-based assessment on classroom practices: Results of the 1996-1997 RAND survey of Kentucky teachers of mathematics and writing*. (CSE Tech. Rep. No. 482). Los Angeles: University of California, National Center for Research on Evaluation, Standards, and Student Testing (CRESST).

Texas Education Agency. (2002). *2002 comprehensive annual report on Texas public schools*. Austin: Texas Education Agency.

## APPENDIX: MATERIALS USED IN ADDITION TO CITED REFERENCES

- Allen, L. H., A. (1999). *Beyond finger pointing and test scores*. Chicago, IL: Cross-City Campaign for Urban for School Reform.
- Ananda, S., & Rabinowitz, S. (2001). *High-stakes and assessment innovation: A negative correlation?* San Francisco: WestEd.
- Ascher, C., Ikeda, K., Fruchter, N. (1998). *Schools on notice: New York state education department's 1996-1997 schools under registration review process*. Institute for Education and Social Policy, New York University.
- Betts, J., & Dannenberg, A. (2002). *School accountability California style: An analysis of recent trends in achievement, school resources and intervention*. Paper presented at the Taking account of accountability: Cambridge, Harvard University.
- California Department of Education. (2002). *Research briefs on student performance: A fresh look at the API score statewide rank (QF01)*. Sacramento, CA: California Department of Education, Policy and Evaluation Division, Research and Analysis Unit.
- Charles A. Dana Center. (1999). *Hope for urban education: A study of nine high-performing, high-poverty, urban elementary schools*. Austin: The Charles A. Dana Center, University of Texas at Austin.
- David, J. (1999). *The road to improvement: Conference report on interventions in low-performing schools*. Frankfort: Partnership for Kentucky Schools.
- Elmore, R. (1990). *Restructuring school: The next generation of educational reform*. San Francisco: Jossey-Bass.
- Fetler, M. (1994). Carrot or stick? How do school performance reports work? *Education Policy Analysis Archives*, 2(13).
- Gray, J. (2000). *Causing concern but improving: A review of schools' experiences*. Research Brief 188. London, GB: Department for Education and Employment.
- Guth, G., Holtzman, D., Schneider, S., Carlos, L., Smith, J., Hayward, G., et al. (1999). *Evaluation of California's standards based accountability system: Final report November 1999*. San Francisco: WestEd.
- Hamman, D., & Schenck, A. (2002). Corrective action and school choice in New York City: An analysis of district funding applications. *Education Policy Analysis Archives*, 10(45).

- Hanushek, E., & Raymond, M. (2002). *Lessons and limits of state accountability systems*. Paper presented at the Taking Account of Accountability. Cambridge: Harvard University.
- Haslam, B. (2002). *Hard work at a fast pace: A study of San Antonio's campus instructional coordinators*. Washington, DC: Policy Studies Associates.
- Hess, G. A., Jr. (1999a). Expectations, opportunity, capacity, and will: The four essential components of Chicago school reform. *Educational Policy*, 13(4), 494-517.
- Hess, G. A., Jr. (1999b). Understanding achievement (and other) changes under Chicago school reform. *Educational Evaluation and Policy Analysis*, 21(1), 67-83.
- Hess, G. A., Jr. (2000). *Changes in student achievement in Illinois and Chicago, 1990-2000* (An occasional paper released by the Brown Center on Education Policy). Washington, DC: The Brookings Institution.
- Just, A., & Boese, L. (2002). *Public school accountability (2000-2001): Immediate Intervention/Underperforming Schools Program (II/USP): How California's low-performing schools are continuing their efforts to improve student achievement*. Sacramento: California Department of Education, Division of Policy and Evaluation.
- Kadamus, J. (1999). *Status report on schools under registration review (SURR)*. New York: New York State Education Department, Office of Elementary, Middle, and Continuing Education (Memo).
- Kadamus, J. (2000). *Status report on schools under registration review (SURR)*. Albany: New York State Education Department, Office of Elementary and Continuing Education (Memo).
- Kadamus, J. (2001). *Report on implementation of a system of accountability for student success (SASS)*. Albany: New York State Education Department (Memo).
- Koretz, D., & Barron, S. (1998). *The validity of gains on the Kentucky Instructional Results Information System (MR-1014-EDU)*. Santa Monica, CA: RAND.
- Ladd, H. (1999). The Dallas school accountability and incentive program: An evaluation of its impacts on student outcomes. *Economics and Education Review*, 18(1), 1-16.
- Ladd, H., & Zelli, A. (2001). *School-based accountability in North Carolina: The responses of school principals* (Working Paper Series SAN01-13). Palo Alto: Terry Sanford Institute of Public Policy Duke University.
- Mintrop, H. (2004). *Schools on probation: How accountability works (and doesn't work)*. New York: Teachers College Press.

- National Association of State Boards of Education. (2002). *From sanctions to solutions: Meeting the needs of low-performing schools*. Alexandria, VA: National Association of State Boards of Education.
- New York City Schools and Community Services. (2000). *The registration review reports: A summary of findings for the 1999-2000 registration review initiative*. Albany: New York State Education Department.
- New York City Schools and Community Services. (2002). *Flash report #7: Year two analyses of performance of extended-time and non-extended time SURR schools*. Albany: New York Public Schools Division of Assessment and Accountability.
- North Carolina Public Schools. (2002). *2000-2001 A report card for the ABCs of public education*. Raleigh: Public Schools of North Carolina. Available from [http://www.ncpublicschools.org/abc\\_results/results\\_01/numberreceiving.html](http://www.ncpublicschools.org/abc_results/results_01/numberreceiving.html)
- North Carolina Public Schools. (2000). *Improving student performance: The role of district-level staff*. Raleigh: Public Schools of North Carolina, Department of Public Instruction.
- Quality Counts 2002. (2002). *Quality Counts 2002: Building Blocks for Success. Education Week, 21(17), 68-169.*
- South Eastern Regional Vision for Education (SERVE). (1997). *Using accountability as a lever for changing the culture of schools: Examining district strategies*. Greensboro, NC: SERVE.
- South Eastern Regional Vision for Education. (2001). *Assisting low-performing schools in the Southeast*. Greensboro, NC: SERVE.
- Texas Education Agency. (1996a). *Statewide Texas educational progress study report No. 1: The development of accountability systems nationwide and in Texas (GE6-601-07)*. Austin: Texas Education Agency.
- Texas Education Agency. (1996b). *Statewide Texas educational progress study report No. 2: Case studies of successful campuses: Responses to a high-stakes state accountability system (GE6-600-04)*. Austin: Texas Education Agency.
- Texas Education Agency. (1997). *Statewide Texas educational progress report No. 9: Expanding the scope of the Texas public school accountability system (GE7 601 07)*. Austin: Texas Education Agency, Office of Policy Planning and Research.
- Texas Education Agency. (2001). *2001 Comprehensive annual report on Texas public schools*. Austin: Texas Education Agency.
- Texas Education Agency. (2002). *2002 Comprehensive annual report on Texas public schools*. Austin: Texas Education Agency.

- Watts, J. (2000). *Getting results with accountability: Rating schools, assisting schools, and improving schools*. Atlanta, GA: Southern Regional Education Board.
- Wong, K. K., Anagnostopoulos, D., & Rutledge, S. (1998). *Accommodation and conflict: The implementation of Chicago's probation and reconstitution policies*. Paper presented at the 1998 Annual Meeting of the Association for Public Policy Analysis and Management in New York.
- Wong, K., & Shen., F. (2001). *Does school district takeover work? Assessing the effectiveness of city and state takeover as a school reform strategy*. Paper presented at the 2001 Annual Meeting of the American Political Science Association in San Francisco.

### Websites Used

For California:

<http://www.cde.ca.gov/psaa/>

For Chicago Public Schools:

<http://accountability.cps.k12.il.us/>

For Kentucky:

<http://www.kde.state.ky.us/KDE/Administrative+Resources/School+Improvement/default.htm>

For North Carolina

<http://www.ncpublicschools.org/abcs/>

For New York:

<http://www.emsc.nysed.gov/nyc/accountability.html>

For Texas:

<http://www.tea.state.tx.us/accountability.html>

For Maryland:

<http://www.msde.state.md.us/Reconstitution/index.htm>