

High-Stakes Accountability, State Oversight, and Educational Equity

HEINRICH MINTROP

University of California, Berkeley

This article argues that outcome-based accountability systems are not likely to close the achievement gap. Using California as an example, it suggests that states need to pay closer attention to learning conditions and powerful programs that deliver quality interventions. The article describes the design of the California low-performing schools program, discusses reasons for its limited effectiveness, and suggests design features that may increase the chances for the state to reach its equity goals. The article argues with an eye on California that input-based design features ought to complement outcome-based accountability. Such designs require standards for adequate learning conditions alongside performance standards, mechanisms to detect systemic performance barriers, and the provision of sophisticated evaluation, high-quality support, and consequences for responsible actors.

In previous decades, the role of federal and state governments in education was limited. On the whole, schools were financed, governed, and supervised locally (Conley, 2003). Testing of student learning was widespread and standardized but rarely used to make decisions about student advancement or measure the performance of schools and educators. Schools were largely autonomous, teachers' work was only minimally inspected, and quality of education was gauged by the provision of inputs to the educational process (Meyer & Rowan, 1978). Federal compensatory programs from the 1960s onward provided resources for additional services to disadvantaged students. Monitoring of these programs was largely limited to compliance reviews of appropriate spending and appropriately targeted services.

This system of local control and loose oversight has dramatically changed due to two developments: (1) In many states, financing of education shifted from local districts to the state as states were compelled to equalize funding across districts, and (2) conceptions of quality shifted from the provision of inputs to the assessment of outcomes (Conley, 2003; Goertz, 2001; Smith & O'Day, 1991; Timar & Kirp, 1988) as dissatisfaction with the low performance levels of schools mounted. This was particularly the case for schools serving traditionally disadvantaged populations where in the minds of

many the slow pace of educational improvement contrasted with increasing costs. Researchers and policy makers proposed to set up accountability systems that were “crystal clear” in their performance standards, attached incentives and sanctions to desired outcomes, and left it largely up to educators to craft strategies that would reach these goals (Hanushek, 1994). Though differing by state, the results of these dynamics were outcome-based or standards-based accountability systems run by states, in some cases on the basis of more equalized funding, that broadly consist of the following elements (Fuhrman, 1999):

- Focus on student achievement, most often measured by simple numerical indicators
- Public reporting of performance indices
- Schools as units of improvement
- Expectations of continuous performance improvement
- Identification and categorization of over- and underperforming schools
- Rewards and sanctions for educators, and promotion and graduation for students, attached to performance

Some systems, such as the ones in North Carolina or Kentucky, added to these elements a fairly elaborate centralized structure for capacity building. Other states, for example Texas or Maryland, relied more on regional or local structures (Mintrop and Papazian, 2003; Council of Chief State School Officers, 2003). Common to these primarily outcome-based accountability systems (perhaps with the exception of New York) is that they do not pay explicit attention to so-called opportunity to learn standards. Although demographic differences in schools’ student intake are sometimes accounted for, for example in California’s “Similar Schools Rank,” indicators of schools’ learning conditions are not imputed in calculating school performance. The federal No Child Left Behind legislation, making outcome-based accountability a national approach to educational equity, has actually taken a step in the direction of opportunity to learn standards by demanding of states to staff schools with “highly qualified teachers.”

THE EQUITY CHALLENGE

The creation of outcome-based accountability systems brought with it a new approach to educational equity. Emphasis shifted from redistribution of resources to incentives as main levers to close the achievement gap. Redistributive policies (Peterson, Rabe, & Wong, 1991) are guided by the

assumption that low performance of schools is substantively a result of under-resourced learning environments. A high frequency of low-performing students in particular schools would be evidence of the schools' needs for additional resources. Incentive policies, on the other hand, place the burden of responsibility for poor performance on employees' work effort. Underperformance becomes associated with failing teachers and administrators. Incentive policies target will over capacity (Thompson & Zeuli, 1999). Some see incentives as a way to prod schools to improve without new expenditures (Hanushek, 1994); others see the need for new expenditures (Fuhrman & Elmore, 2001) but want to make them more effective by coupling capacity building to high-stakes goals. Common to all of these conceptions of primarily outcome-based accountability systems is a belief in the motivational power of goal setting and incentives.

One prime mechanism for equalizing student achievement in outcome-based accountability systems is differential goal setting. While continuous improvement is expected of all schools state-wide, yearly growth increments in lower performing schools must exceed state averages. This is, for example, the approach taken by the states of Kentucky and California where yearly growth expectations for every school are calculated as the difference between a school's baseline in a given year and an envisioned performance ceiling divided by the number of years all schools are given to reach the ceiling. This statistical procedure results in higher growth demands for lower performers. Once state governments have committed to such differential performance goals for equity purposes, it becomes incumbent to design systems that manage to accelerate improvements in the state's lowest-testing schools relative to higher-testing schools.

In this article, I describe one such design, the California low-performing schools program, discuss reasons for its limited effectiveness, and suggest design features that may increase the chances for states to reach their equity goals. I argue with an eye on California that input-based design features ought to complement outcome-based accountability. Such designs require standards for adequate learning conditions alongside performance standards (O'Day & Smith, 1993), mechanisms to detect systemic performance barriers, and the provision of sophisticated evaluation, high quality support, and consequences for responsible actors.

A CASE IN POINT: THE CALIFORNIA LOW-PERFORMING SCHOOLS PROGRAM

As is the case for many other accountability systems, California policies have been in flux. Shifting coalitions in the legislature, fluctuations in the state's fiscal situation, new federal compliance demands and so on have resulted in

a continuous redesign of the accountability system. For several years, since 1999 when the current accountability system was inaugurated, the state relied on two low-performing schools programs as the centerpiece of its accountability system: the Immediate Intervention/Underperforming Schools Program (II/USP) and the High Priority Schools Grant Program (HPSG). The two programs were similar in their core features, but targeted different types of schools.

Other policies and measures beyond the narrow purview of low-performing schools programs are also of importance for the fate of schools lagging behind. Key measures are the certification, recruitment, and retention of quality teachers, building renovation and construction, funds for professional development, and measures related to the instructional program of schools. Naturally, the supply of quality teachers, safe and well-equipped buildings, and effective instructional programs can have a great influence on a school's performance and may be particularly germane for schools performing at the bottom. In California local districts have traditionally been given a decisive role in these areas even though state policies influence local actors. This is particularly true not only for the area of instruction but also for school construction and renovation, often dependent on local bond measures, and for the quality of teachers, often dependent on districts' salary levels and working conditions.

With the advent of outcome-based accountability, low-performing schools programs became the state's prime levers of remedying achievement gaps. Schools performing below state average that did not meet their growth target¹ qualified for the II/USP program. Schools performing in the lowest performance decile independently of previous performance status qualified for the HPSG. Participation was voluntary. Schools selected into the program accepted increased scrutiny and accountability from the state in return for modest, but not trivial funds usable for site capacity building. In other words, California struck a bargain with its schools: higher stakes in return for new funds. Besides additional funds, evaluation and planning were the main intervention levers at the school level.

Identified schools had to contract with an external evaluator that 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 in subsequent years showing evidence of success. The external evaluators negotiated with schools the extent of their fees and services. The state provided to schools a \$50,000 planning grant that could be used to pay the external evaluator and then another \$200 per student each year over 2 years that was to pay for capacity building measures chosen at the school's

discretion. To receive this 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. Remedies listed in the school improvement plan needed to be within the funding parameters of the program. The state envisioned a school to be in the program for 3 years, 1 planning and 2 implementation years, at the end of which the school either exited the program successfully or faced further sanctions. During the 2 implementation years, the school was expected to meet its growth targets in full.

PROGRAM EFFECTS

Entrance and exit rules for the low-performing schools program reflected ambitious equity goals by casting a wide net of eligibility. In II/USP, all below-average schools not meeting their growth targets could procure new funds in return for more intense accountability, extending eligibility to a quarter or so of all schools in the state. The state's far-reaching goal, however, found sharp limitations in the reality of program implementation. To begin with, only about 50 to 60% of eligible low-performing schools in the years 2000 and 2001 were willing to strike the accountability bargain with the state and applied for the program. Thus, almost half of all schools identified as underperforming in 2000 and 2001 chose to bypass the program. The state's funding limitations further curtailed the program. The main program (II/USP) had resources for only 430 schools yearly, a far cry from the number of eligible schools; 935 in 2000; and 1266 in 2001. HPSG funded another 299 schools. The result was that many schools that qualified as underperforming according to the state's definition were not supported by the program. Thus, the extent of the program was limited relative to the state's own educational equity goals and relative to schools' felt needs as indicated by the number of unsuccessful applicants.

In 2003, after 3 years in the program, the first cohort of 430 schools had run through their planning and implementation years. Only about a fourth of the schools qualified for outright exit, having actually achieved the expected growth rates. The state, facing an enormous potential intervention burden, chose to reduce the growth demands put on schools and ended up with merely 24 schools that were subjected to an intermediate process of more intensive support and intervention by a so-called School Assistance and Intervention Team (SAIT) before they may face more serious sanctions. The great majority of the Cohort 1 schools did not sustain the kinds of growth rates over 2 years envisioned by the designers of the system.

In fact, an external evaluation (O'Day & Bitter, 2003) of all three cohorts' performance found that when schools enrolled in II/USP were compared

with a control group of schools that were eligible for the program but did not enroll,² no systematic and statistically significant performance effects could be found for program enrollees in the aggregate. Schools, however, differed widely; some made large gains, while others did not make any improvements. Apparently, the power of accountability incentives and pressures was not strong enough to achieve consistent effects across the system.

In the subsequent sections I argue that this lack of effectiveness may be attributed to three conditions that are inherent in the California design: designers at the state overestimated the power of incentives, underestimated systemic performance barriers and local inequalities in learning conditions, and failed to recognize the need, or lacked the capacity, for designing a tight intervention structure that could deliver support of high quality and intensity.

THE LIMITS OF PRESSURES AND INCENTIVES

It is a hallmark of the type of accountability systems that came into existence in the 1990s and beyond that they are “high stakes.” This means that schools failing to perform according to their state’s (or district’s) expectations face consequences. They 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 with which they were threatened should insufficient performance persist. Many low-performing schools programs foresee sanctions, such as principal and teacher reassignment, state takeover, school reconstitution or closure, and teacher evaluation and dismissal. California is no exception in this regard although the state avoids sanctions with the most stringent consequences for individual teachers.

We have very few research accounts on the responses of schools to pressure and sanctions (Brady, 2003; CPRE, 2001; Hess, 2003; Malen et al., 2002; Hess, 2003; O’Day, in press; O’Day & Bitter, 2003; Wong, Anagnostopoulos, & Rutledge, 1998). But these few accounts suggest that 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.

CREDIBILITY PROBLEMS

The results of more severe sanctions and the implementation of major school redesigns as envisioned by state regulation have shown to be inconclusive (Brady, 2003). For example, 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). Thus, for many teachers, severe sanctions appear impractical and not particularly credible (Mintrop, 2003). Indeed, the centrality of sanctions has faded in many first-generation accountability systems. 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 has been 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 reconstituted three schools, all located in one urban district (Ferguson, 2000). Likewise Maryland, after 5 years of high-stakes accountability, had three schools taken over by the state, schools later assigned to private management organizations. As was mentioned before, in Chicago seven schools were reconstituted in the 1997–1998 school year, but this has not been repeated. Moreover, school principals receive training and support from an "area instructional officer" making the original "probation manager" superfluous. Thus, as a number of first-generation accountability systems have matured, they moved from an emphasis on pressure to one on support to increase their effectiveness.

ACCEPTABILITY PROBLEMS

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. One goal of the systems is to focus teachers on instruction, but attaching too much pressure to a single indicator may have undesirable consequences. Teachers may indeed respond to the pressure by heeding, or perhaps teaching to, the test, but forcing them to narrow the scope of their work creates serious acceptability problems. State assessments may be perceived as invalid indicators of teachers' work; and with reduced educational meaningfulness of the

system, teachers reject the accountability goals as motivating and guiding their work (Mintrop, 2003).

COMMITMENT PROBLEMS

Heightened pressure exacerbates already severe teacher commitment problems in many low-performing schools. Many low-performing schools are not attractive work places, 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 in such schools. Principals under pressure of accountability may act as conduits of pressure creating frenzy which in turn undermines relationships of support and further fragments already unstable organizations. Thus, high pressure may lead to dissatisfaction and exit (Ingersoll, 2001; Mintrop, 2003; Malen et al., 2002). Energy and activism may surge under pressure in the short term, but in the longer term commitment to stay may wane. Findings from Maryland and Kentucky schools (Mintrop, 2003) suggest that higher levels of engagement and effort in response to accountability pressures were not be matched by higher commitment to stay at the pressured school.

CAPACITY PROBLEMS

Research by O'Day and associates found that in Chicago schools' responses to probation bifurcated. Some schools improved rapidly while others lingered in the program. Initial organizational 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, p. 27) responded more favorably. These findings point to the critical importance of capacity building in low-performing schools. It seems that schools with low relational capacity benefit only weakly or not at all from pressure.

Likewise, a study of Maryland schools on probation (Mintrop, 2003, 2004) found that principal leadership, faculty collegiality, cohesion, and trust in the skills of colleagues were stronger in schools that made improvement gains after being identified compared to schools that did not make gains. Neither anxiety about pressure nor acceptability of accountability was higher. If anything, acceptability of accountability was lower in the schools posting gains. At the same time, faculties that were more highly motivated by accountability were not more committed to stay at their negatively labeled schools. By contrast, when site relationships were perceived as strong, commitment rose. Thus, strength of social relations and capacity

rather than the motivational impact of high-stakes incentives and sanctions moved schools towards improvement.

ATTRIBUTION PROBLEMS

Many low-performing schools suffer from low organizational and instructional capacity caused by high teacher turnover, unqualified teachers, unfilled vacancies, lack of instructional materials, weak principals or high principal turnover (CPRE, 2001; Goe, 2001; Hess, 1999; Mintrop, 2004; Reynolds, 1991, 1996). A study of a state-wide database in New York found that lower performing schools tend to be staffed with lower qualified teachers (Lankford, Loeb, & Wykoff, 2002). In the Maryland study, constant teacher turnover, increasing numbers of inexperienced teachers, severe overcrowding, and reassignment of instructional specialists to regular classroom duties due to unfilled vacancies made the achievement of improvement very difficult. At the same time, the studied schools were confronted with educating highly disadvantaged student populations. This in turn brought issues of fairness and attribution to the fore (Mintrop, 2004; O'Day, in press). As schools and teachers felt forced to assume responsibility for critical conditions of student learning over which they lacked control, they rejected accountability altogether. Many teachers who felt unfairly blamed, in turn blamed society and higher levels of the system for schools' performance problems. In this case accountability systems become de-motivating.

Thus, serious credibility, acceptability, commitment, and attribution problems as well as capacity deficits in underperforming schools make it less likely to remedy achievement gaps by merely raising the stakes for educators. It seems that many schools lack baseline stability without which they could hardly be considered responsible actors being in the position to enact a strategy of continuous improvement. Rather, these problems behoove us to thoroughly rethink the wisdom of steering schools with largely outcome-based accountability systems that keep tight reins on performance goals and work motivation via information, incentives, and pressures but leave it up to local responsibility to develop remedies for performance problems.

SYSTEMATIC PERFORMANCE BARRIERS

If what ails schools could be remedied through school-internal action and within the funding parameters of the low-performing schools program, the design of the California program might have worked better. But the reality for California's low-performing schools is different. The data strongly suggest that low performance is strongly associated with systemic conditions of educational disadvantage that the California design is not able to capture,

let alone remedy. To begin with, low-performance status is associated with higher poverty levels, higher percentages of English language learners, and overcrowding (i.e., year-round schools). Moreover, identified low performing schools (II/USP in the state's language) are often clustered in districts that serve traditionally disadvantaged populations and that struggle with challenging conditions. Our own analysis identified 67 districts among the 1,000 or so districts in the state (excluding districts with fewer than ten schools) that had at least 20% of their schools performing below the 20th percentile on the state's academic performance index, or at least 30% of their schools II/USP eligible in 2000 or 2001. We found 4 districts in 2000 that had at least 50% of their schools II/USP eligible, and 14 in 2001; 26 districts in 2000 and 34 in 2001 had at least 30% II/USP eligible schools. In many of these districts proportions of year-round schools and percentages of teachers with emergency credentials were substantially higher than the state average, in a few districts exceeding 50 percent (state average 12%). This points to serious systemic performance barriers (Mintrop, 2002).

An analysis of the content of 65 action plans, written by II/USP schools with the help of independent external evaluators, similarly suggests the inadequacy of the school-internal lens in understanding performance. When schools listed their performance barriers as they were directed to do by state guidelines, they almost all listed school-external factors, such as inadequate facilities, inadequate district curricula and policies, as well as lack of quality teaching personnel, in addition to school-internal barriers (Mintrop, 2002). Thus, the state's own databases suggest the presence of systemic barriers that require remedies beyond the narrow parameters of the low-performing schools program. But, paradoxically, remedies for these systemic barriers were to be school-based and had to be confined within the limits of the state's per pupil funding received by individual schools. Plans that went beyond this frame would be returned to schools for further editing. In other words, the program discouraged remedies for performance barriers that could not be fixed within the financial frame of \$200 per student.

The guidelines for schools participating in HPSG (the state's program for rock-bottom performers) with regard to teacher quality suggest a curious approach that is illustrative of the inadequacy of the school-internal lens: Schools were to describe "what action [they will] take to reduce the number of uncredentialed and inexperienced teachers to at least the district average" (<http://www.cde.ca.gov/iiusp/hpsg/actionplan.html>, p. 7). Apparently the designers of the program were aware that teacher inexperience was a serious problem at these rock-bottom performers. It is understandable that schools *internally* cannot be expected to increase the proportion of experienced teachers beyond what the district context allows, but it is astounding, given that some districts had up to 50% of their teachers uncredentialed, that the "district average" would do as a solution to the

school's problem. Thus, it appears that the California low-performing schools programs were designed with a built-in mismatch between identified systemic performance barriers and allowable solutions.

INADEQUACY OF LOOSE OVERSIGHT AND PROGRAM MANAGEMENT

External evaluation is a key intervention that was required in the California low-performing schools program. The way the state handled this feature illustrates the state's approach. Either policy makers' outcome-based bias made the looseness of the program acceptable or loose management is indicative of severe limitations of state capacity. Neither did the state seize the opportunity to collect systematic data from schools on performance barriers through the program's evaluation feature nor was a management structure put in place that could ensure high-quality evaluations for school-internal consumption and follow-up with targeted support.

External evaluators were chosen through a written application process in which candidates stated their past record of success in improving schools. They were not specifically trained for their evaluation work (External Evaluators Application, California Department of Education, 1999, <http://www.cde.ca.gov/i/iusp>). Keeping within the school-internal lens of the program, their role was apparently one of facilitating internal school change. Data collection on performance barriers was done for the school's own consumption, rather than for the purpose of providing the state with pertinent information on systemic performance barriers.

Since the state department of education had little information on the quality of external evaluators' work, the quality of the intervention in schools often hinged on the uneven quality of specific external evaluators (Goe, 2001; Just et al., 2001). Since the state had neither standards for school learning conditions nor a standardized process of school evaluation in which evaluators could be trained, contracted consultants by default had to rely on professional "connoisseurship" (Wilson, 1996). Connoisseurship evolves through long experience in the educational system and the ability to form realistic expectations based on comparative experiences across many different school contexts. It is very difficult to establish the presence of such connoisseurship among external evaluators by way of a superficial application process, as was used by the state department.

The school action plan appears to have been the key document for the disbursement of implementation funds when the low-performing schools program began. But the state's evaluation of the plans was merely formal and procedural, rather than substantive (State Board of Education minutes, April 12, 2000). The state department did not further process the

information contained in the action plans. It did not conduct summary analyses nor did it compile a summary report that could facilitate the identification of school level factors, or district and state policy-relevant barriers that went beyond school control. Nor did the state department of education conduct analyses of other data that could address systemic district or state barriers to student performance. It is ironic that the state department exerted pressure on schools to make “data-driven” decisions (see Guidelines for Developing an Action Plan, p. 4) when it itself did not make systematic use of data that 1,290 identified low-performing schools with the help of state-financed external evaluators had collected for review and enumerated in their action plans. The effect was that valuable policy-relevant information both with regard to performance barriers and the scope of remedies was lost. Had the state processed this information it would have become clear that the low-performing schools program would have had to reach beyond pressures, incentives, and school grants to become effective.

In sum, while the state held schools accountable through a primarily outcome-based system, it ignored the mismatch between systemic performance barriers and school internal solutions allowable within the program’s parameters, and it forfeited the opportunity to collect quality data on systemic and school-specific factors of underperformance. This information could have enhanced the motivational impact of accountability on school actors by clarifying attribution, and it could have had strong informative potential for policy makers and the public. Furthermore, the state lacked management structures that could ensure quality evaluations and interventions in schools. Program looseness and inattention to systemic performance barriers made it doubtful that the state would be able to achieve its equity goals. Moreover, the continuous paring down of the program’s scale from high numbers of eligible schools to very low numbers of schools receiving intensive support put the state’s commitment to its ambitious equity goals in doubt.

MISSING PIECES OF AN EQUITY-FOCUSED ACCOUNTABILITY SYSTEM

California not only relied on incentives and pressures but backed its equity goals with additional grants to low-performing schools. Schools could use these funds for school-internal capacity building. Yet despite these grants, schools did not improve at higher rates than similar schools that did not receive these funds. A mere grant-making approach does not seem to be a powerful lever of capacity building.

I argue that a different approach may be more promising. The system is in need of standards for adequate learning conditions that accompany

performance standards (O'Day & Smith, 1993), mechanisms to detect systemic performance barriers and unequal learning conditions, and policies that lead to the provision of sophisticated evaluation, high quality support, and forceful intervention. With these elements, the state would be able to steer schools with more fine-grained indicators, recognize root causes of performance problems, design appropriate policies, and design a tighter program of support and intervention.

THE POWER OF OPPORTUNITY TO LEARN STANDARDS

The case of California shows that without adequacy standards for educational inputs, learning conditions and educational practices can diverge widely across the state and can potentially deteriorate substantially without notice in some schools and districts. An example of this is the almost complete 20-year lapse of new school construction in the Los Angeles district, which led to severe overcrowding. Had such standards been in place, schools, parents, and community advocacy organizations would have had criteria on which to evaluate the adequacy of local school operations and lobby their districts and the state for improvements. Likewise, the state would be in the position to gauge if districts were doing their job in providing these conditions or if conditions prevailed that demanded state policy adjustments.

Opportunity to learn standards are the first step in discerning the shared responsibilities of school, district, and state actors for student achievement. Identifying those responsibilities will avert the current unproductive blame game in which the state places the onus of performance on schools, and educators, feeling solely "blamed" by their state for performance problems, in turn "blame" external forces. The motivational power of incentives and pressures suffers when schools are held accountable for causes of underperformance that are in fact controlled by higher levels of the educational system. Outcome-based accountability systems, as currently designed in the United States, are quick to classify "low-performing schools" and "low-performing districts" but fail to acknowledge conditions under which low performance is associated with unequal learning conditions. Opportunity to learn standards make it possible to create a "high need" classification for schools and districts, alongside the low-performance label, that could trigger mandatory provision of support on the part of districts or the state.

DETECTING SHORTCOMINGS AND LOCATING THE PROBLEM

If indeed organizational capacity is a key factor in explaining schools' success in low-performing schools programs, then external reviews need to be conducted based on standards that establish not only adequate levels of

funding but also availability of instructional material, decent facilities, teacher qualifications, stability of faculty, competence of school administration, and so on, all of which are indispensable in stabilizing schools as responsible organizational actors. Including these school quality indicators into the accountability system, in conjunction with multiple indicators of student achievement and behavior, would improve the current unsatisfactory situation in California where schools are evaluated on a narrow set of indicators.

Skillful external evaluators are best suited to perform such comprehensive reviews. They need to be carefully selected and trained in the application of performance and opportunity to learn standards. Data information systems need to be developed that allow the state to monitor learning conditions in lagging schools and districts. Professionally trained personnel are needed to augment evaluation with professional advice and provide support targeted to low-performing and high-needs schools. It seems likely that at present California lacks organizations and personnel who could conduct such sophisticated evaluations and provide targeted support on the scale needed by the system. But this should not deter the state to set clear and ambitious goals and develop a time line for the development of a high-quality school improvement infrastructure. Albeit lacking clear opportunity to learn standards, the state of Kentucky has developed such an infrastructure as part of its accountability system (Mintrop & Papazian, 2003). But the Kentucky low-performing schools program operates on a much smaller scale compared to what would be needed in California.

The problem of underperformance in the state is not restricted to some isolated cases but is widespread and from all indications systemic. Our data have shown that district policies and actions are prominently displayed as problems for schools, but districts also play a key role in solving schools' performance problems (CPRE, 2001). It is therefore indispensable for the state to develop mechanisms that hold districts directly accountable for their schools' performance. When more than a third, or even more than half, of all schools in a given district are not meeting growth targets or when overall district performance ranks well below the state average, state intervention is indicated. Naturally, when schools labor under faulty district policies intervention in district affairs is potentially more powerful than interventions in many schools in one district. When districts do not have the capacity to attract good teachers, build and maintain adequate facilities, or issue coherent policies, they need help from the state in the form of resources and authoritative guidance. Compared to school interventions, district interventions are easier and more difficult at the same time. They are easier because they touch on local policy making, resource allocation, and administration. In these areas, state departments of education have some

expertise, as opposed to the area of teaching and learning where bureaucracies are notoriously ill-suited and the profession ought to reign supreme. District interventions are more difficult because districts have power to resist and marshal political forces whereas schools tend to comply with authoritative measures, though often with minimum enthusiasm.

TIGHTENING THE SYSTEM WITH A NEW SCHOOL IMPROVEMENT INFRASTRUCTURE

The state demands schools and districts tackle performance problems with vigor, ambition, and “no excuses.” It should likewise follow suit and not shrink from the task of building a new institutional framework for school improvement. Ultimately, an agency is needed that develops, systematizes, and oversees external evaluations, interventions, and support for sizable numbers of schools and districts. Such an agency need not actually carry out school or district interventions, but it should identify objective indicators for essential inputs, collect data, monitor conditions in high-need and low-performing schools and districts, and develop model interventions. Given limited resources, this agency ought to concentrate its efforts on schools and districts with serious performance deficiencies.

It is important that this agency be independent from the direct line of authority within the educational system running from the state through districts to local schools. The agency must be able to facilitate reciprocal accountability. It ought to be chartered with a primary commitment to protecting the interests of children in low performing and high-need schools and districts. As an independently chartered agency, it would be able to uncover shortcomings of the state, districts, local schools, and teacher training institutions. This independent agency would organize reviews according to state laws and regulations, mobilize the profession for educational improvements, and make recommendations on required resources, policies, and supports for capacity building. Identified shortcomings on any level should trigger mandatory remedial action on the level that is most appropriate for a solution.

The agency would train external evaluators who can distinguish between site-internal, district and state barriers of performance and identify systemic problems related to districts and the state. The work of evaluation would focus on discovery of “improvement potential” rather than judgment as in the case of the English inspectorate (Grubb, 2000; Wilcox & Gray, 1996). Schools would be evaluated, but subsequently supported in the implementation of their improvement strategies, either by the agency itself or by suitable consultants. Thus the primary role of the agency is the integration of performance data with data on learning conditions, recruitment and training of evaluators, quality assurance of contract work, and the

development of model interventions targeted to high-needs and low-performing schools.

Authoritative reports would be compiled that inform district and state policy makers and the public, not unlike the reports compiled by the English Office for Standards in Education (Ofsted, 2000, at <http://www.ofsted.gov.uk/>). These reports have addressed educators' accomplishments and shortcomings, but also lack of educational provisions that fall under the responsibility of local and national policy makers. For example, in the 2001 report to Parliament, Her Majesty's Chief Inspector of Schools comments that "attention to the teaching of literacy and numeracy is essential" but warns of an "unacceptable narrowing of the curriculum pupils receive" in some primary schools (Ofsted, 2002, p. 1). Or regarding closing the performance gap, the Chief Inspector observes:

Although at primary level, the gap between the highest and lowest performing schools is narrowing, at secondary level, despite improvements within the lowest performing schools, the gap has widened. Many schools facing challenging circumstances are working very hard to improve their performance, but the multiplicity of challenges they face are not all within their power to tackle. (Ofsted, 2002, p. 2)

Parent organizations and advocacy groups could use such authoritative reports to gain information, press for the improvement of local schools, and lobby for local and state policies.

In recent years, states have created accountability systems that rearranged the business of public education. But as illustrated in the California case, the job is incomplete. For the sake of equity goals, a bold step in institution building is needed that mobilizes local policy makers, the teaching profession, and advocacy and community groups.

Notes

1 The first cohort qualified without this criterion.

2 Eligible schools were randomly selected by the state for enrollment in the program.

References

- Brady, R. (2003). *Can failing schools be fixed?* Washington, DC: Thomas Fordham Foundation.
- Conley, D. (2003). *The new landscape of educational policy and governance*. New York: Teachers College Press.
- Consortium for Policy Research in 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: Author.

- Council of Chief State School Officers. (2003). *State support to low-performing schools*. Washington, DC: Author.
- Ferguson, C. (2000). *The progress of education in Texas*. Austin: Southwest Educational Development Laboratory.
- Fuhrman, S., & Elmore, R. (2001). Holding schools accountable: Is it working? *Phi Delta Kappan*, 83(1), 67–72.
- Fuhrman, S. H. (1999). *The new accountability* (RB-27-January 1999). Philadelphia: Consortium for Policy Research in Education.
- Goe, L. (2001). *Implementation of California's Immediate Intervention/Underperforming Schools Program: Preliminary findings*. Paper presented at the Annual American Educational Research Association Conference, Seattle.
- Goertz, M. E. (2001). Redefining government roles in an era of standards-based reform. *Phi Delta Kappan*, 83(1), 62–66.
- Grubb, W. N. (2000). Opening classrooms and improving teaching: Lessons from school inspections in England. *Teachers College Record*, 102(4), 696–723.
- Hanushek, E. A. (1994). *Making schools work: Improving performance and controlling costs*. Washington, DC: Brookings Institution.
- Hess, G. A., Jr. (2003). Reconstitution three years later. *Education and Urban Society*, 35(3), 82–105.
- Ingersoll, R. (2001). Teacher turnover and teacher shortages: An organizational analysis. *American Educational Research Journal*, 38(3), 499–534.
- Just, A. E., Boese, L. E., Burkhardt, R., Carstens, L. J., Devine, 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: California Department of Education, Division of Policy and Evaluation.
- Lankford, H., Loeb, S., & Wyckoff, J. (2002). Teacher sorting and the plight of urban schools: A descriptive analysis. *Educational Evaluation and Policy Analysis*, 24(1), 37–62.
- Malen, B., Croninger, R., Muncey, D., & Redmond-Jones, D. (2002). Reconstitution schools: "Testing" the "theory of action". *Education Evaluation and Policy Analysis*, 24(2), 113–132.
- Meyer, J., & Rowan, B. (1978). The structure of educational organizations. In M. W. Meyer (Ed.), *Environments and organizations* (pp. 78–109). San Francisco: Jossey-Bass.
- Mintrop, H. (2002). *State oversight and the improvement of low-performing schools in California* (Expert witness report for Eliezer Williams et al. versus State of California). Retrieved August 15, 2004, from www.decentsschools.org
- Mintrop, H. (2003). The limits of sanctions in low-performing schools. *Education Policy Analysis Archives*, 11(3). Retrieved September 28, 2004, from <http://epaa.asu.edu>
- Mintrop, H. (2004). *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*. Washington, DC: Office of Educational Research and Improvement, U.S. Department of Education.
- Mintrop, H., & Papazian, R. (2003). *Systemic strategies to improve low-performing schools*. Washington, DC: National Governors Association.
- 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. A., & Smith, M. S. (1993). Systemic reform and educational opportunity. In S. H. Fuhrman (Ed.), *Designing coherent education policy: Improving the system* (pp. 250–312). San Francisco: Jossey-Bass.
- Ofsted. (2002). *The annual report of Her Majesty's Chief Inspector of Schools*. London: Office for Standards in Education.

- Peterson, P., Rabe, B., & Wong, K. (1991). The maturation of redistributive programs. In A. Odden (Ed.), *Education policy implementation* (pp. 65–80). Albany: State University of New York Press.
- Reynolds, D. (1991). Changing ineffective schools. In M. Ainscow (Ed.), *Effective schools for all*. London: David Fulton.
- Reynolds, D. (1996). Turning round ineffective schools. In D. Gray, D. Reynolds, C. Fitz-Gibbon & D. Jesson (Eds.), *Merging traditions: The future of research on school effectiveness and school improvement*. London: Cassell.
- Smith, M. S., & O'Day, J. (1991). Systemic school reform. In S. H. Fuhrman & B. Malen (Eds.), *The politics of curriculum and testing: The 1990 yearbook of the politics of education* (pp. 233–267). London: Falmer Press.
- Thompson, C., & Zeuli, J. (1999). The frame and the tapestry: Standards-based reform and professional development. In L. Darling-Hammond & G. Sykes (Eds.), *Teaching as the learning profession: Handbook of policy and practice* (pp. 341–375). San Francisco: Jossey-Bass.
- Timar, T. B., & Kirp, D. L. (1988). *Managing educational excellence*. New York: Falmer Press.
- Wilcox, B., & Gray, J. (1996). *Inspecting schools: Holding schools to account and helping schools to improve*. Buckingham, UK: Open University Press.
- Wilson, T. A. (1996). *Reaching for a better standard: English school inspection and the dilemma of accountability for American schools*. New York: Teachers College Press.
- Wong, K., & Anagnostopoulos, D. (1998). Can integrated governance reconstruct teaching? Lessons learned from two low-performing Chicago high schools. In R. Macpherson (Ed.), *The politics of accountability* (pp. 26–42). Thousand Oaks, CA: Corwin.
- Wong, K. K., Anagnostopoulos, D., & Rutledge, S. (1998). *Accommodation and conflict: The implementation of Chicago's probation and reconstitution policies*. Paper presented at the Annual Meeting of the Association for Public Policy Analysis and Management, New York.

HEINRICH MINTROP is an associate professor at the University of California, Berkeley. He conducts research on the politics and practice of school improvement. His recent emphasis has been the study of standards-based reform and accountability systems in the United States and Europe. He is the author of the book *Schools on Probation: How Accountability Works (And Doesn't Work)*, recently published by Teachers College Press.