

12-1-2005

Education Policy Analysis Archives 13/48

Arizona State University

University of South Florida

Follow this and additional works at: http://scholarcommons.usf.edu/coedu_pub

 Part of the [Education Commons](#)

Scholar Commons Citation

Arizona State University and University of South Florida, "Education Policy Analysis Archives 13/48 " (2005). *College of Education Publications*. Paper 583.

http://scholarcommons.usf.edu/coedu_pub/583

This Article is brought to you for free and open access by the College of Education at Scholar Commons. It has been accepted for inclusion in College of Education Publications by an authorized administrator of Scholar Commons. For more information, please contact scholarcommons@usf.edu.

EDUCATION POLICY ANALYSIS ARCHIVES

A peer-reviewed scholarly journal

Editor: Sherman Dorn

College of Education

University of South Florida

Volume 13 Number 48

December 1, 2005

ISSN 1068-2341

Corrective Action in Low Performing Schools: Lessons for NCLB Implementation from First-generation Accountability Systems¹

Heinrich Mintrop
University of California, Berkeley

Tina Trujillo
University of California, Los Angeles

Citation: Mintrop, H. & Trujillo, T.M. (2005). Corrective action in low performing schools: Lessons for NCLB implementation from first-generation accountability systems. *Education Policy Analysis Archives*, 13(48). Retrieved [date] from <http://epaa.asu.edu/epaa/v13n48/>.

Abstract

This paper explores what lessons we can learn from the experiences of states that instituted NCLB-like accountability systems prior to 2001 (here called first-generation accountability systems). We looked at the experiences of three smaller states (Kentucky, Maryland, North Carolina), four larger ones (California, Florida, New York, Texas), and two large districts (Chicago and Philadelphia). We analyzed evaluative reports and policy documents as well as interviews with state officials and researchers. We condensed the material into eight lessons: sanctions are not the fallback solution; no single strategy has been universally successful; staging should be handled with flexibility; intensive capacity building is necessary; a comprehensive set of strategies seems promising; relationship-building needs to

¹ This article is based on two CRESST Technical Reports (Mintrop & Papazian, 2003; Mintrop & Trujillo, 2004).



Readers are free to copy, display, and distribute this article, as long as the work is attributed to the author(s) and **Education Policy Analysis Archives**, it is distributed for non-commercial purposes only, and no alteration or transformation is made in the work. More details of this Creative Commons license are available at <http://creativecommons.org/licenses/by-nc-nd/2.5/>. All other uses must be approved by the author(s) or **EPAA**. **EPAA** is published jointly by the Colleges of Education at Arizona State University and the University of South Florida. Articles are indexed by H.W. Wilson & Co. Accepted under the editorship of Sherman Dorn. Send commentary to Casey Cobb (casey.cobb@uconn.edu) and errata notes to Sherman Dorn (epaa-editor@shermamdorn.com).

complement powerful programs; competence reduces conflict; and strong state commitment is needed to create system capacity.

Keywords: corrective action, accountability, No Child Left Behind.

Introduction

According to NCLB, states are to create accountability systems by formulating standards, testing students regularly, defining a baseline, and setting a level of proficiency from 2001 performance levels. Schools are required to attain *adequate yearly progress* (AYP) towards proficiency. AYP can vary from year to year, but all schools need to have reached proficiency for 100 percent of their students by the school year 2013–14. 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 and 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. When they fail to make district AYP for two consecutive years, they enter the improvement stage that primarily entails programmatic changes. After another two years of missing AYP, they are subject to corrective action that may severely curtail their authority.

This paper concentrates on the stage of corrective action and further restructuring. We summarize what lessons might be gleaned from first-generation accountability systems for this stage. Under NCLB, states and districts may soon face the burden of increasing numbers of schools that failed to improve under the softer touch of probation and school improvement. For some states, the NCLB three-stage approach to low-performing schools is novel. But other state governments acted prior to federal legislation. Some jurisdictions identified quite a substantial number of low performing schools, and some states have moved on to more forceful interventions in schools and districts. Although most of these earlier first-generation high-stakes systems echo the structures of NCLB in its basic format, they differ widely in their repercussions for identified low performing schools and districts (Rudo, 2001). States implementing NCLB or aligning their existing accountability system to NCLB can learn from these variations. Insights from first-generation systems can help avoid less promising design features or suggest likely trajectories for certain system designs.

The Research

We looked at three smaller states (Kentucky, Maryland, North Carolina) and four larger ones (California, Florida, New York, Texas). These seven states constitute the main body of our research.

We also looked at Chicago's and Philadelphia's approach to low-performing schools. We selected these systems for five reasons: they are first-generation systems that have spearheaded high stakes accountability in the U.S.; have been in existence for some time; have figured prominently in the public discussion on high stakes accountability prior to NCLB; have gained experiences with corrective action and school redesign; and are covered by some research material. Not all five criteria applied to all jurisdictions.

We asked about the following issues associated with accountability: the kinds of initiatives, programs, or policies undertaken with regard to schools with persistently low test scores that failed the first stage of intervention; the scale on which these programs operated; which set of actors (teachers, school administrators, districts) were the recipients of interventions; how pressures and sanctions were used; what kind of capacity building was provided; what management structures states or districts used for the provision of these services; what evidence of success that might exist; and what lessons could be gleaned from answers to these questions for states that are in the process of designing corrective action or school redesign programs.

Our data are studies, papers, reports, and information from web sites, and we relied on interviews and personal communication with officials to fill gaps.² Although we now have reports on the impact of high-stakes testing on schools in several states, systematic evaluations of low-performing schools programs are rare, and of corrective action initiatives even more so.³ Our 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 design element. Many factors mediate the influence of a particular state or district policy on school performance, including the local context, the specific mixture of interventions, or the time allotted for improvement. It is even more difficult to assess the effectiveness of a specific program relative to other differently structured programs without a common metric that would allow us to compare in a straightforward way.

Given these limitations, we cannot evaluate states' and districts' corrective action efforts, but we can do more than merely describe design features. We refrained from burdening the reader with too much descriptive information.⁴ Rather, we concentrate on lessons learned. We hope that our overview may help systematize and categorize the states' various strategies and their consequences. In this way, we hope to foster an informed discussion about corrective action and school redesign based on previous experiences.

Commonalities and Differences across Systems

Across the states and districts, the following elements, in varied combinations, are most frequently associated with corrective action and school redesign: school improvement grants,

² We conducted background interviews with individuals from the following organizations: Baltimore City Public School System; California Department of Education; The Center for Urban School Policy; The Charles A. Dana Center; Florida Department of Education; Johns Hopkins University; Kentucky Department of Education; Maryland State Department of Education; New York State Department of Education; North Carolina Department of Public Instruction; NYU's Institute for Education and Social Policy; Research for Action in Philadelphia; School District of Philadelphia; Texas A & M University; Texas Education Agency; and WestEd.

³ For examples of evaluations of high stakes testing policies, see Herman (2004); Koretz and Barron (1998); Stecher, Barron, Chun, and Ross (2000); and Stecher, Barron, Kaganoff, and Goodwin (1998).

⁴ For a good descriptive report, see Council of Chief State School Officers (2003).

professional development, new instructional materials, programmatic prescriptions (e.g., pacing plans, structured reading and math programs), new or extension of existing services (e.g., summer school, extended day, after-school), on-site instructional specialists, evaluations or audits, intervention teams or individual change agents, bureaucratic pressures (e.g., reassignment of teachers, principals, external monitors, increased oversight), market pressures (vouchers, school choice, student reassignment, magnet schools), school reorganizations or reconstitutions, teacher recruitment incentives, teacher quality policies, school construction and repair programs, and changes of governance and authority (e.g., special districts, educational management organizations, charters, school takeover, district takeover).

Although NCLB creates some uniformity in states' approaches to low performance by demanding adequate yearly progress towards a proficiency ceiling, the rigor of performance demands and intervention burdens differ across states. These differences influence the chances of persistently low-testing schools to improve and for corrective action and redesign to be successful. Some systems put high demands on schools by either testing student achievement with cognitively complex tests or by expecting growth that was set according to an ambitious performance ceiling. Others took a more moderated approach. They used, 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. Some systems only entered schools into the low performing schools program that were rock-bottom performers, others identified schools on various absolute performance levels that missed their growth targets. Programs differed on what kind of growth it took for a school to exit the program and to shed the low performance label. Moreover, some accountability systems had implemented vigorous district accountability, others had not (Mintrop & Papazian, 2003).

These mechanisms produce low performing schools programs with different improvement challenges and on different scales. These differences also entail varying numbers of schools that fail the first stage of school improvement and are in need of further corrective action. Programs with relatively 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 keep their operational scale low. Generally speaking, 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), first generation accountability systems kept the scale of their programs fairly modest (between 2 and 4 percent), California being the exception.

Table 1

Differences between percent of students scoring proficient on NAEP and state tests (2003)

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

Note: Figures from *Education Week*, Jan. 6, 2005.

NCLB leaves it up to the states to define test rigor and proficiency levels, though the federal NAEP (National Assessment of Educational Progress) tests function as a benchmark that states are to strive for. Tables 1 and 2 show how testing rigor fundamentally structures a state's challenge and intervention burden.

Table 2
Numbers of schools in need of improvement based on AYP (2003–04)

Statistic	CALIFORNIA	TEXAS	KENTUCKY
N	1,626	199	130
Percent	~20%	~6%	~12%

Note: Figures from *Education Week*, Jan. 6, 2005.

States, such as California, with high testing rigor in combination with challenging demographic conditions produce an enormous intervention burden while states with less rigorous tests and more lenient definitions of proficiency, such as Texas, face a relatively modest challenge. Kentucky is a state with medium testing rigor and correspondingly a medium intervention burden. The following lessons apply to states with performance goals in at least the medium range.

Lessons Learned

Although we lack research or evaluation reports about schools under corrective action or redesign that warrant definitive claims as to the effectiveness of particular strategies or designs, we can nevertheless glean a number of lessons, cautionary in nature, from the various states and districts we analyzed.

Sanctions and Increasing Pressures 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. Such systems, relatively undeveloped in the area of support and capacity building, unduly relied on the power of sanctions as fallback solutions. 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 first-generation systems examined here, though each system's mix may differ from NCLB. In California, principals and teachers were threatened to be reassigned. Schools could be taken over by the state. They could be reorganized, closed, or assigned to the management of another educational or non-profit institution. Parents could select a different public school or apply for charter school status ("PSAA," 1999). State takeover was the most severe sanction in the Maryland system (MSDE, 2001). Public hearings, appointment of a special on-site monitor or master, and eventual school closure were envisaged by the Texas regulations as sanctions ("PSSA," 1995). Assignment of an instructional officer, external partner, removal of the principal, and school reconstitution (i.e. staff reassignment and reorganization) figured prominently in the Chicago system (Hess, 2003). Redesign and closure were also primary sanctions in the New York SURR program (Brady, 2003; NYSED, 2002b). Kentucky and North Carolina added penalties to this list that touch individual teachers more severely (Holdzkom, 2001; Ladd & Zelli, 2001; SERVE, 2001). Teachers in low performing schools were evaluated and could be required to take a general knowledge competency test in North Carolina (Manzo, 1998); in Kentucky, as well, they could be evaluated with the possibility of transfer, demotion, or dismissal (David, Coe, & Kannapel, 2003).

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” (David et al., 2003). 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 (David, Kannapel, & McDiarmid, 2000). Actual imposition of final sanctions has been a negligible feature in Kentucky.

In Texas, more severe sanctions akin to the level of corrective action 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 (TEA, 2002). The state has reconstituted only a handful of schools (Ferguson, 2000). Texas primarily relies on the threat of bad publicity to motivate districts and schools to improve performance (Izumi & Evers, 2002; Skrla, Scheurich, Johnson, & Koschoreck, 2003). Likewise in Maryland, after five years of high stakes accountability, the state finally took over four schools and assigned them to private management organizations (MSDE, 2001; CCSSO, 2003).

In New York and Chicago, more severe sanctions played a greater role. Within New York’s Schools Under Registration Review (SURR) program, affecting primarily New York City, some 35 schools have been closed since the inception of the program (NYSED, 2002a). In Chicago, 7 high schools were reconstituted in the 1997/98 school year, but this has not been repeated (Hess, 2003). Moreover, school principals are now receiving training and support from an *area instructional officer* making the original *probation manager* superfluous.⁵

When the present California accountability system was designed, the turn from pressure to support that earlier accountability systems seemed to have undergone was evident. The California program already began with voluntary participation of qualifying schools, though in actuality most schools were ‘volunteered’ by their districts (O’Day & Bitter, 2003). Schools selected into the program accepted increased scrutiny and accountability from the state in return for funds usable for capacity building at the site (Posnick-Goodwin, 2003). 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 and Bitter (2003) as well as data collection by the author. When fewer schools than envisioned met their growth targets, the state refrained from building up pressure. It readjusted growth expectations and added additional intervention layers preceding more severe sanctions. In this way, out of the first cohort of 430 schools accepted into the program, the state identified merely 24 schools that required this additional intermediate intervention when only about a fourth met the state’s original performance demands.

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 one plausible explanation, but other research suggests that, political costs notwithstanding, the pressure strategy is a double-edged sword and not as promising as perhaps originally perceived. The few that are available speak to a number of reasons:

The results of more severe sanctions and the implementation of major school redesigns as envisioned by state regulation have shown to be inconclusive (ECS, 2002, p. 6).

⁵ For a detailed description of Chicago’s recent incorporation of the Area Instructional Officers into the district accountability system, see Chicago Public Schools (2002) available at: http://edplan.cps.k12.il.us/pdfs/cps_education_plan.pdf.

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. Forcing teachers to severely narrow the scope of their work creates serious acceptability problems for the state assessments. As a result, the educational meaningfulness of accountability systems among pressured teachers is low, and teachers reject the system as an intrinsic motivator for their work (Mintrop, 2003).

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 turn-over 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 (Mintrop, 2004, p. 66).

Identifying low performing schools has put the spotlight on glaring capacity deficits in these schools that a motivation strategy alone cannot remedy. This in turn brings issues of fairness and attribution to the fore. When schools and teachers feel forced to assume responsibility 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. In this case accountability becomes counterproductive and de-motivating (Malen, Croninger, Muncey, & Jones, 2002, p. 120).

In sum, in order for accountability systems to work as proper incentive systems, they must appeal to the “better parts” of the profession. High performing teachers and administrators in low-performing schools, in existence in most low-performing schools, are indispensable for a successful reform strategy, and such personnel ought to consider accountability demands as a lever to pull lower performing teachers along. But unduly intensifying pressures and sanctions tend to create defensiveness and turn off the very people on whose willingness, if not idealism, states and districts need to rely.

Thus, in their majority, first generation states have either rarely used or turned away from high pressure as a main lever to motivate teachers. Instead they came to emphasize mild pressure as a means to motivate educators to improve performance. By contrast, 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 too much on the power of pressures to get the job done, that is, not to depend on sanctions as a fallback solution. Rather, states need to construct powerful low performing schools programs that make corrective action and school redesign an uncommon occurrence. Such programs 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 (Mintrop & Papazian, 2003). Ultimately, such systems need to appeal to the values of “the better parts of the profession.”

No Single Strategy Has Been Universally Successful

A number of strategies have been tried for corrective action and school redesign, but evidence shows that their effect is far from conclusive (Brady, 2003).

Reconstitution. In California, previously locally reconstituted schools in the city of San Francisco showed up again on the state’s low-performing schools list and one is actually slated for corrective action again (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, although a number of schools also benefited from the fresh start (Malen et al., 2002). Results from Chicago’s reconstitutions were inconclusive as well. Fundamentally, staff replacements were not necessarily of higher quality than the original teaching staff, and in many schools teacher morale plummeted (Hess, 2003). In New York’s SURR program corrective action and redesign were used more vigorously. Almost fifty schools were reconstituted (ECS, 2002). More than a tenth of the schools were closed. Some schools benefited, yet only about half (153) of the SURR schools have exited the program successfully so far (Brady, 2003; NYSED, 2003).

Educational management organizations. Maryland took over four schools from the Baltimore City school district and passed them on to two educational management organizations (EMOs) (MSDE, 2001c). Under one of the EMOs, only one of its three schools saw consistent gains, one performed unevenly, and one was not improving. In Philadelphia, we have higher numbers of schools that were taken over. One fourth of all district schools were taken over, with 46 managed by different external management organizations and 21 by the district’s newly created Office of Restructured Schools. Here, each provider offers different models of intervention (Travers, 2003a). Preliminary data suggest that the quality and content of the interventions may differ substantially and that the schools managed by the district’s own Office of Restructured Schools may have achieved greater gains than schools managed by EMOs (Useem, 2005). Takeover by EMOs coincided with soaring resignations and teacher turnover in affected schools (Neild & Spiridakis, 2002; Neild, Useem, Travers, & Lesnick, 2003). It also resulted in miscommunication and in some cases overwhelm by principals who felt like they were serving two masters—the EMO and the central office (Blanc, 2003; Bulkley, Mundell, & Riffer, 2004). Thus, takeover by management companies has helped in some cases, but is not universally positive.⁶

A recent multi-state study by the Brown Center at the Brookings Institution finds that schools taken over by EMOs (and run as charter schools) tend to score much lower than their district-administered counterparts, but outscore regular public schools on test score gains (The Brown Center, 2003). The authors suggest that EMO charters’ low test scores are explained by the fact that EMOs tend to take over the lowest performing schools, but the data are not conclusive on this point.

External partners. This feature was widely used in Chicago where each school on probation (i.e. still in the improvement stage) was assigned an external partner (Hess, 2003). Originally, external partners developed their own models of intervention, but disparities in the quality of services concerned the district (O’Day & Finnigan, 2003). In time, the district came to place stronger emphasis on reading, forcing external partners to adapt their work in the schools to meet these literacy goals. Analysts stated that some partners added superficial reading strategies to their

⁶ For an analysis of one EMO’s uneven results, see Bracey (2002).

intervention. This compromised their original model and made them less effective. On the other hand, in reconstituted schools (i.e. those undergoing corrective action) about half of the teaching force found their external partner useful in formulating a shared vision and offering new techniques and strategies after having worked with them for a number of years (Hess, 2003). But an inherent problem in external partner models is the lack of focus on state or district goals and the uneven quality of provided consultant services.⁷

Charters. While the research base on charter schools is expanding, little is known about charter school conversion as a means of corrective action and school redesign.⁸ Available data seem to suggest that converting district-administered schools into charter schools has had uneven results. A multi-state study by the Brown Center on American Education shows that generally charter schools lag behind, or are similar to, regular public schools in absolute performance and gains from year to year (The Brown Center, 2003). Charter schools also tend to show up on states' lists of failing schools in larger proportions than regular public schools. Schools that are converted from district-administered status to charter status are an exception. In the Brown Center study, conversion charters scored more highly than their public school counterparts and start-up charters. The authors point out, however, that conversion charters tend not to be corrective action schools, but schools that are let go by their districts as a form of reward for solid performance, although solid data on this point are missing. Early anecdotal evidence from Philadelphia suggests that charter school conversion without the benefit of an external provider model may be the least successful conversion of the ones tried there.⁹

District takeovers. State takeovers of entire districts have also produced uneven outcomes. Financial management is often cited as the most promising area for potential success by states (Garland, 2003). For example, in Newark, New Jersey, the state reorganized the district and reallocated \$26 million geared toward instruction. When the state stepped into Chicago Public Schools, an anticipated \$4 billion deficit was eliminated. However, equally dramatic *academic* success has been much harder to achieve (Ziebarth, 2002). Academic gains have been mixed at best, most often occurring only after multiple years of intervention. Takeovers in Logan County, West Virginia, Compton, California, and Chicago, Illinois, are heralded as exceptions that yielded some positive academic gains (Garland, 2003).

In a survey of takeover experiences, Garland (2003) details the early lessons about this last resort for low-performing districts: more effective takeovers focus on areas that the state has the capacity to influence, such as financial management, eliminating nepotism, or facilities improvement; attending to the political elements of takeovers through collaboration, negotiation, and local alliances can minimize conflict and resistance; and additional funding, coupled with comprehensive capacity building efforts for both teachers and administrators, can yield more positive results. Nevertheless, he cautions state actors to avoid authoritarian approaches to takeovers and to be mindful of the powerful racial, legal, and political issues that typically accompany these measures.

Former Compton and current Oakland, California, state administrator, Randolph Ward (2004), advocates a comprehensive approach to improving the academic and financial conditions of schools or districts in crisis based on lessons learned during his tenure. These strategies include: developing innovative initiatives for aggressive teacher recruitment and program development;

⁷ Studies of Comprehensive School Reform Design implementation have found analogous disparities in the quality of services provided. See Murphy and Datnow (2002) for examples of such trends.

⁸ For a comprehensive review of the recent research on charter schools, see Bulkley and Wohlstetter (2003).

⁹ In Philadelphia, charter conversions were part of the remedies for low performing schools, not high performing ones, such as those referenced by The Brown Center (2003).

implementing safety net programs like Reading Recovery; creating motivational attendance programs; organizing accelerated learning programs like full-day kindergarten; providing an extended school year; and aligning curriculum with standards-based testing requirements.

Vouchers. Probably the best known example of vouchers attached to low-performance is the state of Florida where students in schools that repeatedly receive an F for their performance can attend private schools on a state voucher. The effectiveness of vouchers as a means to increase competition for low-performing schools is debated. Greene (2001) evaluated Florida's A+ program and found that low performing schools improve more when they face a challenge from vouchers. However, that research has been criticized on methodological grounds (Camilli & Bulkeley, 2001). Thus, at present we do not have sufficient evidence on vouchers as a corrective action strategy.

Intervention teams. These are teams that enter schools as authoritative interveners. They are charged to evaluate schools, prescribe remedies, and help with implementation. In North Carolina, these teams were said to be rather successful (Ladd & Zelli, 2001); in California they worked with mixed success, encountering much resistance at the school level (Posnick-Goodwin, 2003). The two states differ with regard to both operational principles and context. The North Carolina teams were recruited by the state from the ranks of seasoned practitioners and closely worked with schools on an almost daily basis. As teachers in North Carolina cannot engage in collective bargaining, teacher unions are less of a force. In California, the teams were either third-party providers or county offices of education that traditionally were not involved in the day-to-day affairs of regular district schools. They were required to be at the schools a minimum of only three times per year (CDE, 2003). Their initial intervention was tightly circumscribed and, according to interviews with School Assistance and Intervention Team members, tended to eschew instruction.

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

Staging Should be Handled with Flexibility

Although NCLB lays out a straightforward three-stage approach, with corrective action and school redesign being the second or third steps, respectively, schools that are persistently unable to meet AYP are not virgin reform territory for the most part. Many persistently low-performing schools are not stable in their stagnation, but volatile and continuously reconstituting in an unplanned way. Teacher and administrator turnover is often high, external consultants plentiful and ever changing, and district intervention intensified (Mintrop, 2004; Neild & Spiridakis, 2002; Neild et al., 2003). In all likelihood, many low-performing schools, unable to meet federal AYP, will have previously been subjected to substantial local reform measures. Districts that anticipate state action and carry out local school restructuring often move principals and staff, conduct inspections, and mandate programs before a school appears on the state or federal radar screen. When that happens, schools may have to repeat improvement stages or cycles once they enter federal or state corrective action.

Moreover, a comparison of state systems shows how blurred the lines between the stages are in practice. In North Carolina, Kentucky or Florida, the first stage of intervention is already so intense that it could classify as corrective action.¹⁰ By contrast, California's persistently low-testing

¹⁰ For in-depth descriptions of these systems, see Holdzkom (2001) and SERVE (2001).

schools do not even encounter this kind of intensity in the second stage of intervention when they are visited by a state assistance and intervention team (“PSAA,” 1999). Kentucky and North Carolina do not seem to carry out a significantly different corrective action stage. Maryland apparently moved schools from the first stage of local improvement directly into the third stage of takeover and governance change (MSDE, 2001). Something similar has happened in Philadelphia where a fairly large number of the lowest performing schools will make their journey through the NCLB stages as already redesigned schools (Travers, 2003b). As was pointed out above, charter schools tend to show up on states’ failing schools lists in larger proportions than regular public schools. For these schools as well, fundamental redesign happened before school improvement intervention.

In other words, rather than being distinct stages of intervention intensity, NCLB interventions will increasingly look like a *deja vu* to affected schools unless states design intervention approaches that are truly different from all the other things a school has already tried. Such approaches need to decrease turbulence, rather than add to it. Thus, instead of rigid staging, states and districts need flexibility in designing measures that are appropriate to the developmental needs of a given school, an approach that Texas seems to favor.

Intensive Capacity Building Is Necessary

Different approaches to capacity building across states’ low-performing schools programs can inform the design of powerful interventions for the corrective action stage. Generally, state strategies consist of the following elements:

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.

First-generation accountability systems differed in the degree to which these school level oversight and support services were developed. Capacity building intersected with testing rigor, with consequences for schools’ performance and the effectiveness of low-performing schools programs. We distinguish among four patterns:

Ambitious state goals without a capacity building strategy. The toughest challenge ahead was created by the Maryland system in the 1990s. 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. One district, however, was burdened with managing about half of its schools as identified low performers. The state did not develop an elaborate capacity building structure. State monitors were the eyes and ears of the state. Their role in internal school improvement efforts was minimal. Very few low-performing schools managed to exit the program; and indeed schools state-wide stagnated until the system was abandoned. In the Maryland case, state performance demands were decoupled from existing capacities and with a lack of compensatory capacity building, pressures became ineffective or counterproductive.

Less ambitious and flexible goals pegged to state intervention (and teacher) capacity. Texas took an approach that contrasted with that of Maryland in testing rigor, but seemed to exhibit similarities regarding capacity building. The state pegged performance demands at levels that were within reach for most schools and kept the intervention burden relatively small. In 1995, the system identified 267 low-performing schools. The numbers dropped to 59 in 1998, and rose again continuously to 150 in 2002 (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 and tended to exit most identified schools after a short while in the program.

Texas had a decentralized form of governing schools. Most decisions were made on site, and because the state had only limited capacity to support ailing schools, it was indirectly involved in providing assistance (Ferguson, 2000). However, the state required low-performing schools and districts to compile a school improvement plan. It sent peer review teams to schools and districts that visited a school or district for varying lengths of time depending on size of school or district. These peer review teams were made up of state department staff and evaluators that received training with the help of a CD. In addition, the state organized educational support centers that offered their services to low-performing schools and districts, but not exclusively so. Other schools in need of support could contact these centers as well. Texas did not furnish additional monetary grants to low-performing schools. Only a small number of schools were under more direct supervision. As was mentioned above, in the year 2003 only seven schools were visited by monitors and two schools supervised by so called masters. Texas, however, had strong mechanisms built into its accountability system that identified low-performing districts directly and threatened them with further sanctions.

Thus in Texas, relatively small proportions of schools were identified as low performing. Demands on schools were modest and exit criteria within reach. At the same time, the state's support and intervention system was relatively limited. Given that performance demands were more closely pegged to existing teacher capacities (because the system challenged schools in the bottom 20 to 40 percent of the performance distribution with cognitively simple tests), the state could bank on a pressure strategy that succeeded by motivating schools to harvest the low-hanging fruit (Mintrop, 2003).

Figure 1 illustrates (for performance in 8th grade Reading state-wide) the difference between the stagnation of school performance in the Maryland system and the upward trend in the Texas system (leaving the issue of exclusion rates aside).¹¹ Both systems elected to refrain from elaborate capacity-building features within their low-performing schools programs, but in the Texas case upward trends were probably a result of increased pressures around minimum-competency

¹¹ For an analysis of the impact of exclusion rates on Texas achievement patterns, specifically the increased rates of failure of students in grade 9 and the increased number of students leaving school before high-school graduation, see Haney (2000).

standards, a strategy that was pre-empted in the Maryland case by the gulf between school capacity and high state demands.

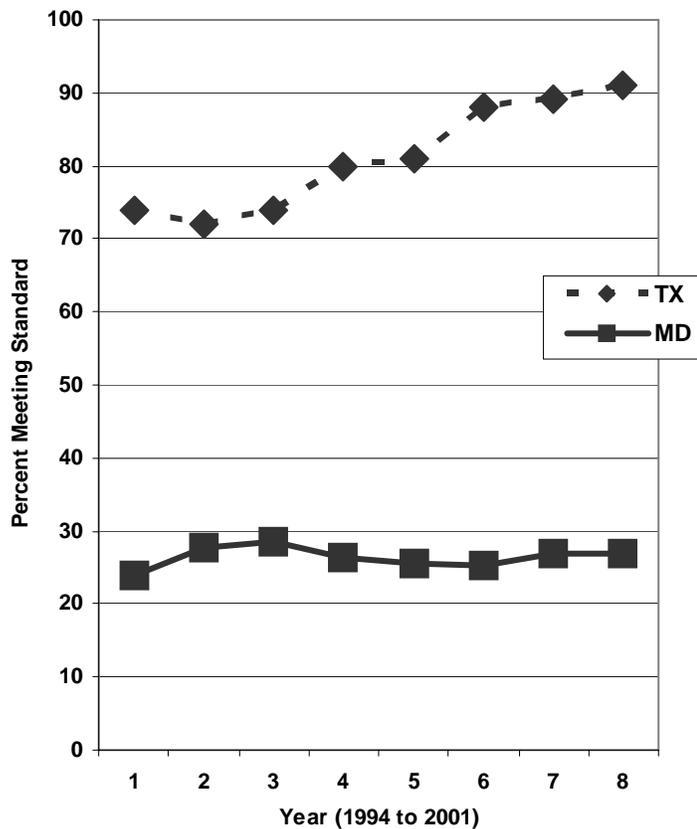


Figure 1
Trends in percent of students meeting standard in Texas and Maryland: Grade 8 reading¹²

Ambitious state goals, substantial school site grants without focused intervention. California’s 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 programs enrolled about 20% of all schools that received an Academic Performance Index,¹³ or about 1,500 schools. The scale of the programs was curtailed by its voluntary feature. Being voluntary, schools and districts decided whether they would apply for additional funds in return for scrutiny and threats of further interventions. In 2001, only 527 or 56% of the 935 eligible schools (for the main program) applied.¹⁴ 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 bypass the program each year. The state ended up accepting only 430 schools each year for funding

¹² From Linn, Baker, and Betebenner (2002).

¹³ California’s Academic Performance Index (API) is a numeric index (or scale) that ranges from 200 to 1000. A school’s API score is one indicator of a school’s performance level. The statewide API target for all schools is 800. A school’s growth is measured by how well it is moving toward or past that goal.

¹⁴ There were actually two programs, the main one applying to all schools below the 50th percentile, the secondary one targeting only Decile 1 schools.

(through its main program). Had all eligible schools been designated, the scale of the program would have been enormous. For capacity building the state relied on the massive disbursement of grants that were attached to a very loosely constructed oversight structure (Mintrop, 2002).

Identified schools had to contract with an external evaluator who was chosen from a state-approved list. Educational reform projects, consultants, county offices of education and later even district offices themselves could apply to this list. 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 negotiated with schools the extent of their fees and services. The state provided schools with a \$50,000 planning grant that could be used to pay the external evaluator, and then another \$ 200 per student per year over 2 years that was to pay for capacity building measures chosen at the school's discretion. During these 2 years, the school was expected to have met its growth targets. 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. Thus, in the California case, the state department kept a low profile. It relied primarily on grant making at a magnitude far greater than most other states we examined, on the capacity of local vendors, the willingness of local districts, and the wisdom of schools to spend the money wisely. A management structure facilitating quality assurance of the support system was only weakly developed. Reports showed that schools' responses to the program varied widely and depended on the varying quality of external evaluators (CDE, 2001; Goe, 2001). A systematic evaluation of the program matching schools enrolled in the low-performing schools program with eligible schools that did not enroll showed no effect on test scores for the enrolled schools (O'Day & Bitter, 2003). Increased accountability pressures in conjunction with substantial grants, relative to other states, did not move these schools on a more successful improvement trajectory than low-performing schools that did not receive this treatment. Qualitative data suggest that the schools lacked sustained quality support and intervention.

Medium rigor, small intervention burden with a developed capacity building structure.

Kentucky is an example of a state that designed its accountability system around performance-based tests with high cognitive complexity. In the 1996 to '98 biennium, the second biennium of the systems' existence, Kentucky entered 250 schools into the low-performing schools program (Cibulka & Lindle, 2001). 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. (Their exit coincided with a redesign of the system (KDE, 2000a), making judgments of effectiveness difficult). In the 2002 accountability cycle, the state identified merely 90 schools as low performing or about 7.5% of the total (KDE, 2000b). Only one third of those were required to accept state intervention which in Kentucky's case was intensive.

Compared to Kentucky, the North Carolina system, with growth expectations pegged to average state growth, yielded a smaller number of identified low-performing schools from its inception. When the state began its ABC tests in the 1996/'97 school year, 123 K–8 schools were identified (7.5% of total). A year later, that number was reduced to only 15 low-performing K–8 schools (0.9%). In subsequent years, the numbers remained low, though they rose again to 44 schools in the 1999–2000 school year, with high schools now being included. But this still constituted no more than about 2% of all schools (NCDPI, 2002). Thus, the North Carolina situation was characterized by a relatively light load of low-performing schools that was consistently held low. Nevertheless, the state's support structure was intensive.

Of the state programs we surveyed, Kentucky and North Carolina had fairly elaborate systems in place that provided oversight and support to schools under direct supervision from the

state department. Services were sustained over one school year or longer, and specifically targeted to low-performing schools achieving state goals. As part of the state's support for its schools *in need of assistance*, Kentucky provided modest additional school improvement funds. In the 2002–2003 year, \$2 million was budgeted for the 90 schools. For example, elementary school grants ranged from \$12,000–\$38,000 per biennium.

A school inspection was conducted by state-sponsored Scholastic Audit Teams, which included a Highly Skilled Educator (HSE), a teacher, a principal or other administrator, a parent, and a university-based educator (KDE, 2000a). The audit teams were trained for their task. The audit teams visited each school for about a week. Once the scholastic audit was conducted, schools used the results to write their school improvement plans. The lowest category of performers (Level III) received mandated assistance from an HSE for the entire biennium; the others received voluntary assistance. School plans were written with the help of the designated HSE and were submitted to the state department for review and approval. A state-certified person other than the HSE also conducted an evaluation of school personnel at all Level III schools. Principals at all three levels were required to participate in staff development to enhance leadership skills.

HSEs had to demonstrate prior ability to bring about high levels of student performance and went through a rigorous hiring and training process. Each HSE received two weeks of training and follow-up training at quarterly meetings. Mentors from the state department provided assistance in problem solving and support to HSEs. HSEs were expected to serve on-site at least 80% of their work time. Their activities included but were not limited to: staff development, classroom observations of instruction, demonstration lessons, grants writing, tutoring, and creation of model lessons (David et al., 2003; Holdzkom, 2001). In addition, a team of HSEs that specialized in organizational management was formed and could be assigned to more than one school at a time, given the needs of a particular school. In the 2002–2003 school year, there were 52 HSEs 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 et al., 2000; Holdzkom, 2001). A reported key focus of the work of the HSE was curriculum and instructional alignment to the instructionally complex state assessments. 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, 2000), although it is difficult to isolate the impact of HSEs in the whole school environment. Significant challenges for the program were sustaining the change once HSE had left school grounds, creating an appropriate match between the HSE and the school, and maintaining a strong pool of HSEs (David et al., 2000).

In North Carolina, no additional funds were allocated to low-performing schools, but these schools received intensive oversight and support. Low-performing schools were 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 worked with a school for one academic year on a daily basis. The teams' tasks were similar to the ones HSEs in Kentucky carried out. In addition, they reported to the local school board or the state department on the school's progress.

Assistance team members participated in a 4-week comprehensive training in topics similar to those in Kentucky. In addition, the assistance teams could participate in 2 extra weeks of training in a program specifically designed to reduce minority achievement gaps and were encouraged to go to conferences regarding specific subject areas or grade-level content. A team of five people working at the state level provided 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 was instrumental in moving schools on the path to improvement.

Goals, program scale and capacity building strategies. High-quality support and oversight need to be an integral part of a low-performing schools program in both the school improvement and corrective action stages. The need for strong support grows in proportion to performance demands.

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 modestly complex performance demands and a high level of guidance by way of a state core curriculum. But at the same time, 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 very low organizational capacity to begin with, establishes a high exit threshold, and leaves it up to (overburdened) districts to provide necessary capacity building runs into trouble even when it keeps the statewide load of identified schools fairly small. We saw tendencies of such a pattern in the original Maryland program 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; such a design runs into trouble. We saw tendencies of such a pattern in the 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 is an example of a system that 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. The problem with such a system, however, is that it may encourage teaching to a cognitively basic test in highly pressured schools.

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. 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 and focused intervention, especially when coupled with ambitious performance goals. Even low-performing schools programs on a small scale designed for the first stage of school improvement required elaborate capacity building structures. For the corrective action stage, these requirements increase manifold.

A Comprehensive Set of Strategies Seems Promising

If there is one characteristic that stands out from systems that keep the number of low performing schools low and make a consistent difference in their lowest performing schools, it is comprehensiveness. For example, Florida uses a comprehensive approach to corrective action schools that includes professional development, instructional support, work on test preparation, help with assessment, extended school days, and parent in-services.¹⁵ In Kentucky, intervention starts off with a comprehensive week-long scholastic audit based on 9 standards and 88 indicators. Highly Skilled Educators (HSEs) are assigned to schools and expected to be on-site at least 80% of the time during their two years at the school. David et al. found that HSEs' activities are "remarkably similar across the sample schools" (2003, p. 10). They fall into the following categories: professional development, curriculum alignment, classroom instruction, test preparation, leadership, school organization and decision-making, and resource procurement. These categories are similar to those used in the Scholastic Audit. North Carolina uses a similarly encompassing approach. When intervention teams enter the school, they evaluate all educators in the school and can recommend dismissing anyone who does not improve at the end of the year (SERVE, 2001). Among other things, they conduct classroom observations, work closely with principals, conduct model lessons, and streamline budgets. The schools are required to implement the team's corrective actions. Team members participate in a 4-week training on data analysis, cultural diversity, curriculum alignment, teacher performance and evaluation, and team building (Holdzkom, 2001).

The Chancellor's District in New York City, emulated by other inner-city districts, was a similarly comprehensive approach to persistently low-testing schools, but added to the mix a supportive district structure that acted as a surrogate for schools' dysfunctional home district (Snipes, Doolittle, & Herlihy, 2002). Intervention in the special district consisted of the following elements (Phenix, Siegel, Zaltsman, & Fruchter, 2005): reduced class size; extended school day and year; after-school programs; mandated instructional programs, schedules, and curricula; prescribed professional development, with at least four on-site staff developers; extra time; a teacher center and teacher specialist assigned to each school; student assessments; supervisory/district support; restaffing and replacement of most principals and many ineffective teachers; more intense monitoring and mentoring; and incentives for recruiting qualified teachers (e.g., signing bonuses).

Researchers and interviewed program administrators point to two factors that in their minds made a key difference: the special district removed a school from a failing district and put it in a very nurturing one, and a comprehensive set of organizational, curricular, instructional and personnel interventions were given to schools as a bundle, avoiding isolated quick fixes (Phenix et al., 2005). However, even with this intense intervention, preliminary data suggest that Chancellor's District schools achieved only moderate improvement in student performance; only half of the enrolled schools were removed from the state list of low-performing schools; and one-fifth had to be closed. Yet, overall fourth graders in the special district outperformed SURR schools in reading, i.e. those schools enrolled in the state's low-performing schools program, even when controlling for student and school characteristics, teacher resources and per student expenditures (Phenix et al., 2005).

In summary, it appears that comprehensiveness is a key characteristic that makes interventions sufficiently different from all the other things that schools have tried before and that makes corrective action programs effective. Comprehensiveness includes interventions at the school

¹⁵ For more descriptions of Florida's accountability system, see SERVE (2001) and the Florida Department of Education's Bureau of School Improvement website at <http://www.bsi.fsu.edu/>.

and district levels. But even these comprehensive approaches cannot overcome some of the performance barriers that exist in the highest-need and lowest-capacity schools and districts.

Relationship-Building Needs to Complement Powerful Programs

Many low-performing schools are not attractive work places, and under current labor market conditions, low-performing schools are often staffed with lower-skilled teachers and large numbers of new, insufficiently trained teachers with low commitment to stay (Lankford, Loeb, & Wyckoff, 2002). 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 (LeCompte & Dworkin, 1991). Mintrop found that in schools improving in the low-performing schools program principal leadership and faculty collegiality and cohesion as well as trust in the skills of colleagues were stronger (Mintrop, 2004). Bryk and Schneider point to the importance of trust among administrators, teachers, and parents as a key resource for school improvement (Bryk & Schneider, 2002). O'Day (2004) found that initial capacity was a key factor in explaining why some schools improved when targeted by low-performing schools programs and others did not. Elementary schools with higher "peer collaboration, teacher-teacher trust, and collective responsibility for student learning" responded more favorably (p. 26). Creation, or "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 summarizes insights from inspection reports on 900 schools "under special measures," the English equivalent to schools under corrective action (Gray, 2000, p. 20).

Under corrective action, districts and states intervene deeply into the core of a school's operation, often mandating specific programs and prescribing specific operations that can be monitored fairly easily. Implementation of effective programs is desirable and especially necessary when schools are staffed with many insufficiently qualified teachers. Under the pressures of corrective action, however, such implementation raises the specter of compliance, managerial control, and programmatic standardization as the main levers of school improvement. Following the lead of the above cited literature, implementation of powerful programs ought not come at the expense of developing professional norms of high expectations and trusting relationships. Such norms are not only necessary for teachers to collectively assume responsibility for student learning, but are important in fostering and maintaining teacher commitment to stay. Moreover, if 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 interventions that incorporate work on internal organizational norms and building trust may be a good way to improve on that front.

Governance changes, for example the installation of an EMO, are often accompanied by heightened political conflict around new relationships of authority and may lead to a decline in social stability.¹⁶ Redesigns have the potential of actually diminishing a school's social capacity (Malen et al., 2002). Intervention strategies need to compensate for these negative consequences of social disruption. Thus, corrective action and redesign strategies need to create a balanced effect on instructional programs, educators' professional norms of performance, commitment to stay in the low-performing school, and trust among school actors. Such balance is apt to stabilize the low-performing school.

¹⁶ For analyses of how EMO management was associated with the social instability of Philadelphia schools, see Neild and Spiridakis (2002) and Neild et al. (2003).

Competence Reduces Conflict

When schools enter the stage of corrective action, they are no longer able to heal themselves and improve solely based on their own internal strengths. Rather, they are in need of external change agents who can provide new tools, such as programs, coaching, advice, and facilitation. While in the first stage of school improvement, state pressure and the signaling of urgency may increase schools' motivation to marshal their own forces; in the corrective action stage pressure takes a back seat to capacity building. Something essential needs to be added to a school under corrective action that it previously lacked. We argued earlier that this "something" is not an isolated quick fix (e.g., a program, a governance change, a principal change, etc.), but a set of strategies that comprehensively integrates the technical and social layers of the organization.

To provide such a collection of strategies with sufficient high quality requires the careful recruitment of highly skilled intervention personnel. This is a key challenge for all systems we examined. Supply of high quality personnel is a theme that runs through many reports and interviews, regardless of the specific models or structures that are implemented.¹⁷ In California, schools complained about a lack of powerful expertise on the part of the state's new school intervention teams (Posnick-Goodwin, 2003). In Philadelphia, the strength of district or EMO-based restructuring seems to rely on the ability of the entity in charge to recruit skillful and committed educators (principals, teachers, staff developers, instructional specialists, etc.) to the schools. Where they fail to do so, or are not empowered to do so by regulations as in Philadelphia, the effort may be undermined irrespective of the specific governance structure.

The uneven service quality of third-party consultants in a number of systems was already mentioned. North Carolina and Kentucky recruit school practitioners with a track record of leadership in their schools and districts in order to insure proximity to the people that need to be reached by interventions. It has been a challenge for Kentucky to find enough highly qualified candidates for the job of Highly Skilled Educator year after year, as previous cohorts return to their districts. The shortage of educators with these skills is evident in the frequent complaint from districts that HSEs are sorely missed in the district's own operations (David et al., 2003).

When external interveners enter schools without a strong base of competence, problems arise. Schools complain of serving two masters (Blanc, 2003; Bulkley, Mundell, & Riffer, 2004). Traditional lines of authority are more likely to clash with new ones, for example state-empowered external interveners, when new authority is not backed up and legitimized by new ideas, new capacities, or new services that promise to be a benefit to the school.¹⁸ In other words, before states (or districts) decide to send new intervention teams, external partners, EMOs, etc. as authoritative executors of corrective action into persistently low-performing schools, they should be sure about the providers' potential to offer comprehensive services with competence. These services need to make a marked difference in schools that in many instances "have tried it all before."

Strong State Commitment Is Needed to Create System Capacity

Corrective action and school redesign cannot be done on the cheap. We know from first-generation accountability systems that merely mandating new programs, subjecting a school to *zero-*

¹⁷ Some examples of this theme include the experiences in Chicago (Hess, 2003) and Philadelphia (Neild & Spiridakis, 2002; Neild et al., 2003).

¹⁸ For examples where external agents conflict with traditional lines of authority, see Garland (2003) and Posnick-Goodwin (2003).

based staffing as in reconstitution, pairing it up with external consultants, or passing it on to new management will not be sufficient for those persistently low-performing schools that have high needs and low capacity to begin with (Mintrop & Papazian, 2003). Successful states and districts show that highly competent personnel and comprehensive intervention capacity are not readily available and have to be developed over time. Particularly, corrective action in *district* administrations seems to be virgin territory for many states.

As NCLB implementation progresses through the stages of corrective action and school redesign, more schools and districts will have to be targeted and states' efforts need to grow. But recent fiscal problems in many states and districts make a vigorous state effort doubtful (Richard, 2004). Comprehensive programs (for example, New York's Chancellor's District, Kentucky's Highly Skilled Educators program, and North Carolina's Assistance Team program) have seen cuts. Some states have stopped expanding their programs or retrenched (such as California). Poorer states, such as Mississippi or Alabama, have been unable to pay for assistance to all of their highest-need schools, let alone pay for the development of a school improvement infrastructure that NCLB implementation will require. The gap between what is federally required for successful corrective action and redesign and what states are able or willing to offer at this point is large in many instances. New ways of financing systems' capacity for providing comprehensive and highly competent interventions need to be found.

Conclusion

The eight lessons learned from first-generation accountability systems can be condensed into one final lesson for the implementation of NCLB. First-generation attempts have shown that the task of continuous school improvement requires a sophisticated school improvement infrastructure of high quality that comprehensively 'moves on all fronts' and goes beyond incentives, sanctions, and even additional grants for capacity building. Yet, NCLB has magnified the challenge even further. The more stringent corrective action requirements of the law are likely to create larger intervention burdens for states than many of the previous systems examined in this paper.¹⁹ The high-stakes features of the law have been called bold by supporters and draconian by detractors. But compared to the enormous challenges of the task, we conclude from our data on first-generation accountability systems that even the law's presumably rigorous corrective action features tinker on the margins. The corrective action incentives and sanctions no doubt can cause movement among responsible actors on all levels of the system. But if the law is implemented in the tradition of procedural compliance, it will produce much commotion and comparatively little improvement. The enormity of the task at hand requires the federal government, states, districts, and schools to go far beyond NCLB and proactively search for powerful, high quality and comprehensive ways of reform and institution rebuilding.

Alternatively, states could reduce testing rigor or keep rigor down. As demands are pegged to existing capacities, a pressure strategy seems more promising and intensive (and expensive) capacity building expendable. Basic literacy and numeracy may rise in such a system by limiting and committing schools in the lower performing spectrum to cognitively simple instruction (at least in the short or medium time frame). First-generation systems such as the ones in Maryland and California show that ambitious performance goals without a well-structured and well-supported capacity building strategy create ineffective low-performing schools programs with undesirable political consequences. Conversely, a system that banks on pressures to teach to a cognitively simple

¹⁹ For more on this topic, see Mintrop and Papazian (2003).

test may confront us with the undesirable trade-off between the ends of achieving basic literacy and numeracy by means of severely curtailing the spectrum of educational goals. Accountability systems designed in the medium range of cognitive complexity with modest pressures and reasonably elaborate capacity building structures may be a good start.

References

- Blanc, S. (2003, Summer). Supports and challenges in the multiple provider model: Principals offer mixed perspectives on first year of takeover. *Philadelphia Public School Notebook*.
- Bracey, G. (2002). *The market in theory meets the market in practice: The case of Edison Schools*: Education Policy Studies Laboratory: Education Policy Research Unit (EPRU), College of Education, Arizona State University, retrieved November 2, 2005, from <http://www.asu.edu/educ/epsl/EPRU/documents/EPRU%202002-107/EPSTL-0202-107-EPRU.htm>.
- Brady, R. (2003). *Can failing schools be fixed?* Washington, DC: Thomas B. Fordham Foundation.
- Bryk, A.S., & Schneider, B.L. (2002). *Trust in schools: A core resource for improvement*. New York: Russell Sage Foundation.
- Bulkley, K., Mundell, L., & Riffer, M. (2004). *Contracting out schools: The first year of the Philadelphia diverse provider model*. Philadelphia, PA: Research for Action.
- Bulkley, K., & Wohlstetter, P. (2003). *Taking account of charter schools: What's happened and what's next?* New York: Teachers College Press.
- California Department of Education. (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: Author.
- California Department of Education. (2003). *School Assistance and Intervention Team (SAIT): Three-year process*. Sacramento, CA: Author.
- Camilli, G., & Bulkley, K. (2001). Critique of 'An Evaluation of the Florida A-Plus Accountability and School Choice Program'. *Education Policy Analysis Archives*, 9(7), retrieved November 2, 2005, from <http://epaa.asu.edu/epaa/v9n7/>.
- Chicago Public Schools. (2002). *Every child, every school: An education plan for the Chicago Public Schools*. Chicago, IL: Author.
- Cibulka, J., & Lindle, J. (2001). *The politics of accountability for school improvement in Kentucky and Maryland* (Tech. Report, Vol. IV). Washington, DC: Office of Educational Research and Improvement, U.S. Department of Education.
- Council of Chief State School Officers. (2003). *State support to low-performing schools*. Washington D.C.: Author.

- David, J., Coe, P., & Kannapel, P. (2003). *Improving low-performing schools: A study of Kentucky's Highly Skilled Educator Program*. Lexington, KY: Partnership for Kentucky Schools.
- David, J., Kannapel, P., & McDiarmid, G.W. (2000). *The influence of Distinguished Educators on school improvement: A study of Kentucky's school intervention program*. Lexington, KY: Partnership for Kentucky Schools.
- Education Commission of the States. (2002). *State interventions in low performing schools and school districts*. Denver, CO: Author.
- Ferguson, C. (2000). *The progress of education in Texas*. Austin, TX: Southwest Educational Development Laboratory.
- Garland, L. (2003). *Navigating treacherous waters: A state takeover handbook*. Lanham, MD: Scarecrow Press.
- Goe, L. (2001, April). *Implementation of California's Immediate Intervention/ Underperforming Schools Program: Preliminary Findings*. Paper presented at the annual meeting of the American Educational Research Association, Seattle, WA.
- Gray, J. (2000). *Causing concern but improving: A review of schools' experiences*. London: Department for Education and Employment.
- Greene, J. (2001). *An evaluation of the Florida A-Plus accountability and school choice program*. New York: The Manhattan Institute.
- Haney, W. M. (2000). The myth of the Texas miracle in education. *Education Policy Analysis Archives*, 8(41), retrieved November 2, 2005, from www.epaa.asu.edu/v8n41/.
- Herman, J. (2004). The effects of testing on instruction. In S. Fuhrman & R. Elmore (Eds.), *Redesigning Accountability Systems for Education* (pp. 141–166). New York: Teachers College Press.
- Hess, G.A. (2003). Reconstitution—three years later: Monitoring the effect of sanctions on Chicago high schools. *Education and Urban Society*, 35(3), 300–327.
- Holdzkom, D. (2001). *Low-performing schools: So you've identified them—now what?* Charleston, W.V.: AEL, Inc.
- Izumi, L., & Evers, W. (2002). State accountability systems. In L. Izumi & W. Evers (Eds.), *School Accountability: An Assessment by the Koret Task Force on K–12 Education* (Vol. Publication No. 512: 2002). Stanford, CA: Hoover Institute Press.
- Kannapel, P., & Coe, P. (2000). Improving schools and school leaders. In R. Pankratz & J. Petrosko (Eds.), *All children can learn: Lessons from the Kentucky reform experience*. San Francisco, CA: Jossey Bass.

- Kentucky Department of Education. (2000a). *Assistance for schools: Guidelines for Scholastic Audit*. Frankfort, KY: Author.
- Kentucky Department of Education. (2000b). *Commonwealth Accountability Testing System: Interim accountability cycle briefing packet (1996–1997 to 1999–2000)*. Frankfort, KY: Author.
- Koretz, D., & Barron, S. (1998). *The validity of gains on the Kentucky Instructional Results Information System*. Santa Monica, CA: RAND.
- Ladd, H.F., & Zelli, A. (2001). *School-based accountability in North Carolina: The responses of school principals*. Durham, NC: Terry Sanford Institute of Public Policy at Duke.
- 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.
- LeCompte, M., & Dworkin, A. (1991). *Giving up on school: Student dropouts and teacher burnouts*. Newbury Park, CA: Corwin Press.
- Malen, B., Croninger, R., Muncey, D., & Jones, D. (2002). Reconstituting schools: “Testing” the “theory of action”. *Educational Evaluation and Policy Analysis*, 24(2), 113–132.
- Manzo, K.K. (1998, June 17). N.C. lawmakers revoke teacher-testing plan. *Education Week*, p. 25.
- Maryland State Department of Education. (2001). *Fact Sheet #5: School Reconstitution: State Intervention Procedures for Schools Not Progressing Toward State Standards*. Baltimore, MD: Author.
- Mintrop, H. (2002). *State oversight and the improvement of low-performing schools in California, expert witness report for Eliezer Williams et al. v. State of California*. San Francisco, CA, retrieved November 2, 2005, from http://128.121.78.24/decentsschools/expert_reports/mintrop_report.pdf.
- Mintrop, H. (2003). The limits of sanctions in low-performing schools: A study of Maryland and Kentucky schools on probation. *Education Policy Analysis Archives*, 11(3), retrieved November 2, 2005, from <http://epaa.asu.edu/epaa/v11n3.html>.
- Mintrop, H. (2004). *Schools on probation: How accountability works (and doesn't work)*. New York: Teachers College Press.
- Mintrop, H., & Papazian, R. (2003, December). *Systemic strategies to improve low-performing schools: Lessons from first-generation accountability systems* (CSE Tech. Rep. No. 617). Los Angeles: National Center for Research on Evaluation, Standards, and Student Testing.
- Mintrop, H., & Trujillo, T. (2004, December). *Corrective action in low-performing schools: Lessons for NCLB implementation from state and district strategies in first-generation accountability*

- systems* (CSE Tech. Rep. No. 641). Los Angeles: National Center for Research on Evaluation, Standards, and Student Testing.
- Murphy, J., & Datnow, A. (2002). *Leadership Lessons from Comprehensive School Reforms*. Thousand Oaks, CA: Corwin Press.
- Neild, R.C., & Spiridakis, K. (2002). *Preliminary results: Philadelphia new teacher survey*. Philadelphia, PA: Research for Action.
- Neild, R.C., Useem, B., Travers, E., & Lesnick, J. (2003). *Once & for all: placing a highly qualified teacher in every Philadelphia classroom*. Philadelphia, PA: Research for Action.
- New York State Education Department, Office of New York City School Improvement and Community Services. (2002a). *Identification and removal of schools from Registration Review: 1989–2002*. Unpublished data, New York City.
- New York State Education Department, Office of New York City School Improvement and Community Services. (2002b). *The Registration Review Process*. New York City: Author.
- New York State Education Department, Office of New York City School Improvement and Community Services. (2003). *2002–03 Registration Review Initiative: A summary of the Registration Review report findings*. New York City: Author.
- North Carolina Department of Public Instruction. (2002). *2000–2001: A report card for the ABCs of public education*. Raleigh, NC: Author.
- O'Day, J. (2004). Complexity, accountability, and school improvement. In S. Fuhrman & R. Elmore (Eds.), *Redesigning accountability systems* (pp. 15–43). 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.
- O'Day, J. & Finnigan, K. (with Wakelyn, D.). (2003). *External support to schools on probation: Getting a leg up?* Chicago, IL: Consortium on Chicago School Research.
- Phenix, D., Siegel, D., Zaltsman, A., & Fruchter, N. (2005). A forced march for failing schools: Lessons from the New York City Chancellor's District. *Education Policy Analysis Archives*, 13(40), retrieved November 2, 2005, from <http://epaa.asu.edu/epaa/v13n40/>.
- Posnick-Goodwin, S. (2003, September). Under close scrutiny. *CTA California Educator*, 8.
- Public School System Accountability, Tex. Ed. Code. § 39.131 (1995).
- Public Schools Accountability Act, Ca. Ed. Code. § 52055.5 (1999).
- Richard, A. (2004, January 7). States unable to help all struggling schools. *Education Week*, pp. 19, 23.

- Rudo, Z. (2001). *Corrective action in low-performing schools and school districts*. Austin, TX: Southwest Educational Development Laboratory.
- SERVE, Inc. (2001). *Assisting low-performing schools in the Southeast: A SERVE special report*. Greensboro, NC: Author.
- Skrla, L., Scheurich, J., Johnson, Jr. J., & Koschoreck, J. (2003). Accountability for equity: Can state policy leverage social justice? In L. Skrla & J. Scheurich (Eds.), *Educational Equity and Accountability: Paradigms, Policies, and Politics*. New York: RutledgeFalmer.
- Snipes, J., Doolittle, F., & Herlihy, C. (2002). *Foundations for success: Case studies of how urban school systems improve student achievement*. Washington, DC: MDRC for the Council of Great City Schools.
- Stecher, B. M., Barron, S. L., Kaganoff, T., & Goodwin, J. (1998, May). *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).
- Stecher, B.M., Barron, S.L., Chun, T., & Ross, K. (2000, August). *The effects of the Washington state education reform on schools and classrooms* (CSE Tech. Rep. No. 525). Los Angeles, CA: National Center for Research on Evaluation Standards and Student Testing.
- Texas Education Agency. (2002). *2002 comprehensive annual report on Texas public schools: A report to the 78th legislature*. Austin, TX: Author.
- The Brown Center. (2003). *Charter schools: Achievement, accountability, and the role of expertise*. Washington, DC: Brookings Institution.
- Travers, E. (2003a). *Characteristics of schools under diverse providers 2002–2003*. Philadelphia, PA: Research for Action.
- Travers, E. (2003b). Philadelphia school reform: Historical roots and reflections on the 2002–2003 school year under state takeover: A talk initially presented at The Friends Association of Higher Education and Friends Council on Education Conference at Swarthmore College and Pendle Hill. *Penn GSE Perspectives on Urban Education*, 2(1), retrieved November 2, 2005, from <http://www.urbanedjournal.org/commentaries/c0007.pdf>.
- Useem, E. (2005). *Learning from Philadelphia's school reform: What do the research findings show so far?* Philadelphia, PA: Research for Action.
- Ward, R. (2004). *Improving achievement in low-performing schools: Key results for school leaders*. Thousand Oaks, CA: Corwin Press.
- Ziebarth, T. (revised 2002). *State takeovers and reconstitutions*. Denver, CO: Education Commission of the States (ECS).

About the Authors

Heinrich Mintrop

University of California, Berkeley

Email: mintrop@berkeley.edu

Heinrich Mintrop taught middle school and high school for over a decade in both the United States and Germany before he entered into his academic career. He received a Ph.D. in Education from Stanford University (1996). He is currently an associate professor of Education at the University of California, Berkeley. As a researcher, he explores issues of school improvement and accountability in both their academic and civic dimensions. He has recently published the book *Schools on Probation: How Accountability Works (and Doesn't Work)*, at Teachers College Press. At UC Berkeley, he is involved in the Principal Leadership Institute that aims to prepare strong leaders for high-need urban schools.

Tina M. Trujillo

University of California, Los Angeles

Email: tina.trujillo@ucla.edu

Tina Trujillo is a doctoral student in Urban Schooling at the University of California, Los Angeles, Graduate School of Education and Information Studies. Her current research interests include systemic district instructional reforms and low-performing schools' responses to high-stakes accountability policies.

EDUCATION POLICY ANALYSIS ARCHIVES <http://epaa.asu.edu>

Editor: Sherman Dorn, University of South Florida

Production Assistant: Chris Murrell, Arizona State University

General questions about appropriateness of topics or particular articles may be addressed to the Editor, Sherman Dorn, epaa-editor@shermamdorn.com.

Editorial Board

Michael W. Apple

University of Wisconsin

Greg Camilli

Rutgers University

Linda Darling-Hammond

Stanford University

Gustavo E. Fischman

Arizona State University

Gene V Glass

Arizona State University

Aimee Howley

Ohio University

William Hunter

University of Ontario Institute of Technology

Daniel Kallós

Umeå University

Thomas Mauhs-Pugh

Green Mountain College

Heinrich Mintrop

University of California, Berkeley

Anthony G. Rud Jr.

Purdue University

Terrence G. Wiley

Arizona State University

David C. Berliner

Arizona State University

Casey Cobb

University of Connecticut

Mark E. Fetler

California Commission on Teacher Credentialing

Richard Garlikov

Birmingham, Alabama

Thomas F. Green

Syracuse University

Craig B. Howley

Appalachia Educational Laboratory

Patricia Fey Jarvis

Seattle, Washington

Benjamin Levin

University of Manitoba

Les McLean

University of Toronto

Michele Moses

Arizona State University

Michael Scriven

Western Michigan University

John Willinsky

University of British Columbia

EDUCATION POLICY ANALYSIS ARCHIVES
English-language Graduate-Student Editorial Board

Noga Admon
New York University

Jessica Allen
University of Colorado

Cheryl Aman
University of British Columbia

Anne Black
University of Connecticut

Marisa Burian-Fitzgerald
Michigan State University

Chad d'Entremont
Teachers College Columbia University

Carol Da Silva
Harvard University

Tara Donahue
Michigan State University

Camille Farrington
University of Illinois Chicago

Chris Frey
Indiana University

Amy Garrett Dikkers
University of Minnesota

Misty Ginicola
Yale University

Jake Gross
Indiana University

Hee Kyung Hong
Loyola University Chicago

Jennifer Lloyd
University of British Columbia

Heather Lord
Yale University

Shereza Mohammed
Florida Atlantic University

Ben Superfine
University of Michigan

John Weathers
University of Pennsylvania

Kyo Yamashiro
University of California Los Angeles

Archivos Analíticos de Políticas Educativas

Associate Editors

Gustavo E. Fischman & Pablo Gentili

Arizona State University & Universidade do Estado do Rio de Janeiro

Founding Associate Editor for Spanish Language (1998—2003)

Roberto Rodríguez Gómez

Editorial Board

Hugo Aboites

Universidad Autónoma
Metropolitana-Xochimilco

Dalila Andrade de Oliveira

Universidade Federal de Minas
Gerais, Belo Horizonte, Brasil

Alejandro Canales

Universidad Nacional Autónoma
de México

Erwin Epstein

Loyola University, Chicago,
Illinois

Rollin Kent

Universidad Autónoma de
Puebla. Puebla, México

Daniel C. Levy

University at Albany, SUNY,
Albany, New York

María Loreto Egaña

Programa Interdisciplinario de
Investigación en Educación

Grover Pango

Foro Latinoamericano de
Políticas Educativas, Perú

Angel Ignacio Pérez Gómez

Universidad de Málaga

Diana Rhoten

Social Science Research Council,
New York, New York

Susan Street

Centro de Investigaciones y
Estudios Superiores en
Antropología Social Occidente,
Guadalajara, México

Antonio Teodoro

Universidade Lusófona Lisboa,

Adrián Acosta

Universidad de Guadalajara
México

Alejandra Birgin

Ministerio de Educación,
Argentina

Ursula Casanova

Arizona State University,
Tempe, Arizona

Mariano Fernández

Enguita Universidad de
Salamanca. España

Walter Kohan

Universidade Estadual do Rio
de Janeiro, Brasil

Nilma Limo Gomes

Universidade Federal de
Minas Gerais, Belo Horizonte

Mariano Narodowski

Universidad Torcuato Di
Tella, Argentina

Vanilda Paiva

Universidade Estadual Do
Rio De Janeiro, Brasil

Mónica Pini

Universidad Nacional de San
Martín, Argentina

José Gimeno Sacristán

Universidad de Valencia,
España

Nelly P. Stromquist

University of Southern
California, Los Angeles,
California

Carlos A. Torres

UCLA

Claudio Almonacid Avila

Universidad Metropolitana de
Ciencias de la Educación, Chile

Teresa Bracho

Centro de Investigación y
Docencia Económica-CIDE

Sigfredo Chiroque

Instituto de Pedagogía Popular,
Perú

Gaudêncio Frigotto

Universidade Estadual do Rio
de Janeiro, Brasil

Roberto Leher

Universidade Estadual do Rio
de Janeiro, Brasil

Pia Lindquist Wong

California State University,
Sacramento, California

Iolanda de Oliveira

Universidade Federal
Fluminense, Brasil

Miguel Pereira

Catedrático Universidad de
Granada, España

Romualdo Portella do

Oliveira

Universidade de São Paulo

Daniel Schugurensky

Ontario Institute for Studies in
Education, Canada

Daniel Suarez

Laboratorio de Políticas
Públicas-Universidad de
Buenos Aires, Argentina

Jurjo Torres Santomé

Universidad de la Coruña,
España