Knowing the Right Thing to Do:

School Improvement and Performance-Based Accountability

Richard F. Elmore
Harvard Graduate School of Education
and Consortium for Policy Research in Education (CPRE)

Acknowledgements

This paper was prepared by Richard F. Elmore of Harvard University and Consortium for Policy Research in Education (CPRE). The NGA Center for Best Practices would also like to thank the following for individuals for their valuable contributions: John Thomasian, director of the Center; Christopher Mazzeo, senior policy analyst; Ilene Berman, deputy director; Dane Linn, director, of the Center's Education Division; and Kathy Skidmore-Williams, publications manager, from NGA's Office of Communications.

Preface

The challenge of turning around low-performing schools raises important questions about how state accountability systems can be deployed to support schools through the improvement process. In this paper, Richard F. Elmore of Harvard University looks in detail at two schools classified as low-performing in their states and draws lessons from the experiences of these schools for state policies regarding school improvement and accountability. According to Professor Elmore, knowing the right thing to do is the central problem of school improvement. Like the two schools portrayed in the paper, many schools classified by states as low performing are actually "improving schools" that have leveled off in performance and now require help from states to continue to improve. These schools differ fundamentally from the stereotypical image of a "failing school" in public and policy discourse.

State accountability systems need to aggressively build teacher, leader and organizational capacity in the state's low-performing schools. To develop accountability systems that can push all schools to get better, the author argues that governors and other state policymakers should take the following actions:

- develop an infrastructure at the state and local levels to provide expertise and support to low-performing schools;
- ensure accountability systems can distinguish among schools that are improving and those that are not;
- require school districts to develop systemwide instructional improvement strategies;
- create incentives to stabilize the population of teachers and administrators in low-performing schools; and
- continue to examine and improve the design of accountability systems.

Introduction

During the past decade, almost every state has initiated some form of performance-based accountability for schools. These accountability systems vary in the detail of their design, but they have several important characteristics in common. They all involve periodic testing of students, reporting of test scores by school, and some mechanism for identifying schools needing improvement. What happens after these schools are identified has been the subject of considerable experimentation at the state level. Some states have tried various forms of state-sponsored intervention (e.g., sending experts into schools to help turn them around). Other states have focused on sanctions, giving these schools a set amount of time to meet performance targets and then closing and reconstituting them after they have failed to meet those targets. The issue of how to handle "low-performing" schools in state accountability policies is one where there is little settled knowledge and a good deal of room for learning.

The advent of No Child Left Behind (NCLB), the most recent reauthorization of the federal Elementary and Secondary School Act, has focused attention in a new way on low-performing schools. The law sets higher expectations for schools and creates clear and immediate sanctions for weak performance. As a result of the law, significant numbers of schools—estimates are several thousand[1]—will likely be identified as "in need of improvement." States and local districts are charged with the responsibility for bringing the performance of schools in line with the expectations of the new law. Yet nothing in the recent history of state accountability efforts has equipped states or localities to handle the number of schools that will likely be classified as low performing under NCLB. More importantly, the problem of low-performing schools raises fundamental questions about the practice of school improvement (e.g., what does it look like when it is actually happening?) and about the design and implementation of accountability systems (e.g., what should be done at the state and local levels to support schools through the improvement process?).

This paper examines the problem of low-performing schools from the inside out, from the perspective of the kinds of schools to which accountability policies are usually directed. In what follows, I focus primarily on two schools, both currently classified as "low performing" under their existing state accountability system, and likely to be classified as "in need of improvement" under NCLB. I look in detail at the process by which these schools are trying to improve and from this analysis derive a working theory of what the process looks like when it is successful. The paper concludes with a discussion of current state accountability systems and recommendations for how governors and other policymakers can recast these systems to better assist low-performing schools in their states.

Thornton and Clemente: Portraits of Two Low-Performing Schools

Consider the following two of the many schools I visited recently that are currently classified by their states as low performing:[2]

Thornton Elementary School

In a second-grade classroom at Thornton Elementary, I observe a teacher with one-third of the students in one corner of the classroom doing guided reading, a form of literacy instruction in which the teacher and students jointly read aloud and discuss a book with an explicit focus on the author's meaning as well as the readers' response to the text. Guided reading is new to the teacher, though she is a veteran of 20 years. She is concentrating very hard. The students are also working very hard, and they seem to be reading successfully and responding to the book. Each student in the class will rotate in one of three groups through guided reading in the course of the literacy block—a 90-minute period every morning devoted to reading and writing. While the teacher is focusing on the 8 students in the guided reading group, the remaining 16 students in the classroom are doing various activities. Two reading specialists are working individually with two students who are obviously struggling readers on specific problems of phonics and word identification. A classroom aide is supervising a group of students that seems to be filling out worksheets. Some students are reading on their own and writing in journals. There are books in considerable

School Accountability Under No Child Left Behind (NCLB)

Under NCLB, states set a minimum percentage of students required to be proficient on state reading and math assessments in every school. These performance targets are called the state "starting points." These targets rise incrementally until all students are expected to be proficient in reading and math in 2013–14. Low-performing schools are required to make larger gains in the percentage of proficient students because the target is further away. Schools that already meet the performance target are required to make smaller gains or, at least initially, no gains at all. In addition, schools must include 95 percent of the entire student body and each subgroup of students in the assessments and must meet one additional indicator (e.g., graduation rate for high schools).

For schools or subgroups that miss state targets, NCLB includes another way for schools to make adequate yearly progress or AYP. "Safe harbor" allows a school or subgroup to meet AYP if the school or subgroup decreases the percentage of nonproficient students by 10 percent. To meet safe harbor, a school or subgroup must also have 95 percent of its students included in state reading/language arts and math assessments, and meet the state's additional academic indicator.

Students attending schools that miss their targets for two straight years must be offered the option of moving to another public school within the school district. Schools must also develop a plan for improvement that promotes parental involvement, uses 10 percent of the school's Title I funds for professional development, and incorporates a teacher mentoring program. After three years of missing the state performance target, parents of children in low-performing schools must be offered the option of using federal Title I dollars to purchase supplemental educational services from an approved provider on the open market. After four years of missing the state performance target, the school becomes the subject of "corrective action." This requires the district to formulate a plan for the school that might include replacing staff, decreasing management authority at the school level, appointing outside experts to advise the school, lengthening the school day or year, or restructuring the school.

The measures required by the law are cumulative—that is, schools in corrective action also are required to provide public school choice and supplemental services. If the school fails to meet its target under corrective action in one year, the district must take one of the following actions:

- the school can be closed and reopened as a charter school;
- the district can replace all or most of the staff relevant to the school's failure;
- the district can be contracted to a private management company;
- the district can be put in state receivership; or
- the district can be subjected to another governance arrangement.

If at any point in this process, the school has two successive years of meeting its target, it ceases to be a failing school. Performance failures, however, are cumulative. That is, the further behind its performance target the school falls, the greater the gains it is required to make in order to meet performance expectations.

quantity available to students. Student writing is prominently displayed on the walls. In general, the classroom appears to be orderly, quiet, and efficiently run. Behavior problems are few. Students seem receptive and relatively happy. The adults seem to be focused, working hard, and highly motivated.

The principal and superintendent have worked out a professional development strategy for the school that focuses time during the school day and during designated professional development days on priority instructional areas. The teachers uniformly say this is the best work they have done. The teaching force at Thornton is a veteran group; the least experienced teacher has been there 12 years.

During the course of the morning, I visit several classrooms at Thornton. To the casual observer, it would be difficult to see why Thornton is a low-performing school. Teachers are working hard. Students are highly engaged. There are extra adults to work with struggling students. The classrooms and hallways are orderly and clean. Thornton certainly looks nothing like the stereotype that laypeople might have about failing schools—chaotic, disorderly classrooms and teachers obviously out of their depth with both content and student discipline. In fact, most observers would probably say that, overall, Thornton represents a strong and positive environment for students. Most observers would be hard-pressed to say what Thornton should be doing differently, or why it is low performing.

The superintendent and the principal have invited me to the school because, after some initial, modest success on the state reading and writing test, Thornton's test scores have gone flat. The student population at Thornton is more than 80 percent poor, with equal numbers of African-American and Hispanic students. The school is in an economically depressed city. The patterns of student performance at Thornton reflect similar patterns in other elementary schools in the district. The state reading and writing test is a challenging exam for even the highest-performing schools in the state. For Thornton, it is daunting. The superintendent and principal report that teachers in the school are demoralized by their designation as a failing school. They feel they have given the school's new literacy program their best shot. They believe they have changed their practices dramatically. The changes they have made are visible in all the classrooms in the school, they feel, and yet they are still not making progress toward the standards of performance they are expected to meet.

Clemente Middle School

Shift now to Clemente Middle School, a school of about 1,000 students, grades six through nine, in a large Northeastern city. Essentially all the students at Clemente meet the income requirement for free or reduced price lunch—the prevailing measure of poverty. They are predominantly Afro-Caribbean, Spanish-speaking, recent immigrants, with a significant number of African-American students. Many students are from families that might be classified as the working poor—they perform the basic services of the economy with low compensation. It is mid-morning, and I am observing a seventh-grade language arts class taught by a novice teacher—a Teach For America corps member, one of several in this school. The teacher is a new graduate of a prestigious Northeastern liberal arts college. She is young, energetic, and highly engaged in her work. She is African-American and has a strong rapport with her students. There are about 15 students in her class today. One student sits off by herself in a corner, focusing on something on her desk—perhaps a "time-out" discipline problem. The rest of the students sit comfortably in moveable desks, focused on the teacher. The lesson covers topic sentences and lead paragraphs—key elements of the state's writing test for middle grades these students will take next year. The teacher is carrying on a lively discussion about topic sentences that students are asked to use as the basis for their writing. As students volunteer ideas and write them down in their notebooks, the teacher actively engages them in a discussion about what they will say and how they will say it.

This classroom is one of a number I have observed this morning, and the patterns are similar: active teachers, highly engaged students, instruction targeted at skills that are useful on their face and, at the same time, included in the state reading and writing test. I do not see a discernible difference between the novice and the experienced teachers

on these dimensions. The principal takes me to visit a couple of classrooms where he knows he has problems with the teachers. These classrooms are noticeably less engaging places for students. The teachers are clearly struggling with the fundamentals of teaching; they are also aware that they are not doing great work. One of the students asks me if I am from the state. Faculty and students know they are in a low-performing school; they are used to being visited, and it does not make them happy.

Clemente has four assistant principals, each with instructional and professional development responsibility for a grade level in the school. The assistant principals are clearly present in classrooms. The principal and assistant principals have a strategy for professional development in key subjects with teachers. While time is limited, teachers participate and say the work is valuable to them in the classroom.

Clemente Middle School is a vibrant and exciting place visually. It is a relatively new building, with a large atrium as a central feature. It has a privately funded arts program in which students produce stunning visual art, and writers from a neighboring city visit while teachers and students conduct author studies of their work. The building exudes energy and student work is visible everywhere.

I am in the school because performance, after a brief gain, has gone flat, well below the target level required to keep the school from being classified as failing. The superintendent and the principal want me to see, hear, and feel what the school is like—not just examine the test scores. The superintendent thinks the principal of the school is one of the best in the district, and the superintendent is worried about losing him to a neighboring district with much higher-performing schools. Teacher turnover in the school and district is about 15 percent per year. About 40 percent of the teachers in the district have four or fewer years of experience. Virtually none of the Teach For America corps members stays after their two-year term is over. The district invests heavily in professional development in literacy and math, but the superintendent says that once the new teachers have received the basic staff development program, they are attractive recruits for neighboring suburban districts that offer much higher salaries.

Again, to most observers, Clemente would not be considered a low-performing school. It has its share of weaker teachers, which in a school of 1,000 is probably a high number. But what you see as you walk the halls and visit classrooms are powerful examples of students doing interesting and creative work, teachers working hard to engage students in learning that is clearly connected to what the state tests measure, and students largely responding in the ways teachers want them to. Take away the discouraging test scores, and you have a school that most observers would say is a decent place for kids to learn.

"State accountability systems generally do not distinguish among schools that are engaged in a long-term improvement process and schools that have not begun such a process."

These schools are not atypical of the legions of schools that will be classified as "in need of improvement" under NCLB. I have been in several such schools in several localities during the last few years, and I would say that Thornton and Clemente are more representative of what I have seen than not. To be sure, I have also been in low-performing schools that more closely resemble the common stereotype of failing schools: schools that show little or no evidence of consistent expectations around the quality of instruction or student performance; schools in which the adults assign responsibility for low student performance to families and communities rather than to themselves; schools in which the resources available to support student learning are managed, if at all, in a chaotic and scattered way; and schools in which teachers and students cannot answer the most basic questions about the purpose and direction of their work.

State accountability systems generally do not distinguish among schools that are engaged in a long-term improvement process (like Thornton and Clemente) and schools that have not begun such a process. Likewise, these systems do not distinguish among schools that have stabilized their internal workings and started on the process of improving

instruction and those that have not met this requirement. I will say more about this later, but it is important to acknowledge the difference—in both policy and practice—between low-performing schools that are improving and those that are not, even though they may look very similar in their student performance.

Internal Accountability and School Improvement

What low-performing schools fundamentally lack is something my colleagues and I have come to call "internal accountability."[3] These schools lack agreement and coherence around expectations for student learning, and they lack the means to influence instructional practice in classrooms in ways that result in student learning. In our research, high internal accountability leads directly to observable gains in student learning. Some low-performing schools lack internal accountability on anything but the most basic expectations—order in the hallways, for example. But, as the Thornton and Clemente examples illustrate, some of these schools actually are engaged in developing internal accountability and have had some success in generating increased student learning, but they are still at risk of not meeting the performance requirements of the new federal law.

I think these two schools can teach us something valuable about the process of improving low-performing schools. I believe that existing policies toward low-performing schools are often based on limited knowledge about school failure. Therefore, policies that are nominally designed to improve these schools may have the opposite effect—they prevent low-performing schools from succeeding. These policies are often based on cultural stereotypes about failing schools, stereotypes that reflect commonly held beliefs and values that are deeply embedded in American culture and political discourse. These falsehoods include believing that schools fail because the people in them—administrators, teachers, and students—don't work hard enough; and that they are lazy, unmotivated, and self-serving. The purpose of policy is to make people use their existing capacities to the maximum—to wake people up and make them work harder.

Schools also fail, in the conventional view, because the people in them lack the appropriate urgency and focus on the essential tasks of learning. Given the opportunity, adults and students in schools will do whatever they want to do, rather than what they are supposed to do. So policies should give schools clear guidance coupled with clear rewards and sanctions that send a clear message about what the work is supposed to be. Finally, the conventional view is that schools fail because they lack the proper incentives to succeed. If schools and the people in them were threatened with closure or loss of enrollment, the argument goes, they would be motivated to discover what they needed to do to succeed. These beliefs are embedded in accountability policies that focus on external rewards and sanctions as motivators for teachers, administrators, and students.

I have spent the better part of my professional career studying the impact of incentives on schools, so I won't argue that incentives don't matter. They do. Nor would I argue that schools can't improve, to a degree, by getting people more focused, encouraging them to make better use of their existing capacities, or making them work harder. One of the main insights I took away from my recent visits to low-performing schools is how clearly most of these schools have gotten the message that the policies are sending them. The problem is that the message doesn't tell them what to do about the problem, other than to "get better." As Thomas Schelling, one of the most astute economic theorists on the subject of incentives, once said, "The problem with most incentive structures is not getting people to do the right thing. It's getting people to know the right thing to do."

"Knowing the right thing to do is the central problem of school improvement."

Knowing the right thing to do is the central problem of school improvement. Holding schools accountable for their performance depends on having people in schools with the knowledge, skill, and judgment to make the improvements that will increase student performance. As Thornton and Clemente illustrate, these improvements are often not obvious even to people who are committed and knowledgeable themselves. Moreover, the process of improvement

is far from smooth, even when the conditions for improvement are relatively supportive. This is particularly apparent when we take a closer look inside Clemente Middle School.

Clemente: A Second Look

This may sound unusual, but the core problem at Clemente was that the teachers were working too hard. Novice teachers—and even some veteran teachers—often equate "good" teaching with teaching that keeps students amused, interested, and seemingly engaged. This means students with eyes forward, paying attention, not causing any discipline problems, and responding in a timely way to the teacher's questions. At Clemente, the "good" teachers in the building had adopted a style in which they were doing virtually all the work in classrooms and the students were doing very little. The teachers felt they were giving it their best shot, and the students were engaged and amused. But when you looked at student work, it was clear that not much was happening. To keep students' attention focused on the front of the room, teachers were asking predominantly factual questions—questions that could be answered literally by the student pulling the information straight out of the text on the desk in front of them. When teachers did ask questions that required higher levels of cognitive demand—interpretation, argument, analysis—the overall pace of previous questions meant that waiting even a short period of time for a student response seemed like ages, so the teacher quickly moved on to the next question before the students could fully engage in the previous one. The actual written work that students were being asked to produce—remember this is a class that prepares students to pass the state writing exam, which includes open-ended writing prompts—was likewise short and truncated, apparently because the teachers judged that the students needed a faster pace and more concrete tasks to stay engaged.

What was happening at Clemente is what often happens in the early stages of instructional improvement—teachers were developing rudimentary norms of practice that signal their collective commitment to student success. They were, in short, developing internal accountability around student learning and performance. Teachers were doing what they thought they should do—working hard, being enthusiastic, demonstrating that they can hold the attention of the students. An observer would think that "good teaching" was going on in a significant number of classrooms and wonder why the results weren't more impressive.

More importantly, teachers were generally doing what they knew how to do. In the absence of careful analysis of the kind of practice that would lead to success on a demanding writing exam they continue to do what they regarded as "good teaching"—and what many would tell them was good teaching—without recognizing that it was precisely that kind of teaching that was producing the disappointing performance. To get markedly different results, teachers would have to learn to do something they didn't know how to do. To do that they need access to skills and knowledge that would help them to understand and enact those practices in their classroom.

Compounding Clemente's problem is its sheer scale: 1,000-plus students, about 60 teachers, 4 assistant principals, and other support personnel. The school has made significant progress in creating a strong and supportive culture around student learning; it is inviting for students and a critical mass of adults clearly care about them. There clearly is a nucleus of teachers who are developing strong norms about what good teaching looks like, even though those norms are not very well aligned with the performance the school is being asked to produce. But the problems of bringing these nascent norms into the classrooms of Clemente are considerably more complex than at a smaller school. All four assistant principals have to understand which teachers need which kinds of development. Each assistant principal has to work with 15 to 20 grade-level teachers, most of them with little experience. Each teacher has an entry point for developing his or her practice and a set of dispositions about the collective expectations of the school. Teacher-by-teacher, the work looks difficult. At this scale, it looks formidable.

Within weeks of visiting Clemente, I visited Buckingham—a middle school in a neighboring state. Buckingham is, by any standard, an impressive and exceptional public school—a high-performing school that can explain how it got that way. Even though it is in an affluent community, it consistently outperforms its peer schools, as does the district in which it resides. The district superintendent is a leading practitioner in his field and skilled at developing talent in others. Buckingham's principal is young and energetic and clearly a competent instructional leader. The district and

the school provide substantial support to its teachers, and exert substantial pressure on them to improve their practice and student learning. The problem at Buckingham was that students leaving the school and moving into the high school were having difficulty meeting the high school's expectations for reading, understanding, and interpreting high-level content. This problem did not show up in the state assessment, but it did show up in the assessments given by the high school teachers. To meet this challenge, Buckingham undertook an ambitious program of reading instruction across the middle school curriculum. Teachers in science, mathematics, history, and social studies, as well as those in English/language arts, were spending some portion of each day working on problems of reading in their content areas. The results of this strategy were disappointing, and the superintendent and principal wanted to know why.

Upon closer look, the problem at Buckingham was similar to Clemente's—the teachers were doing most, if not all, the work. If you watched what the students were doing, you saw them looking at the teacher waiting to be called on, and then going to the appropriate place in the text to retrieve the fact. The high school teachers wanted students who could read high-level text, analyze it, have ideas about it, and articulate those ideas. What middle school instruction was doing—and with ruthless efficiency—was teaching students how to answer factual questions in class.

I give this example to illustrate that Clemente and Buckingham are both improving schools. They are essentially working on the same problem, albeit under vastly different conditions and for vastly different stakes. Buckingham flies under the radar, because, as in most affluent communities, most of its students are at or above the required level of performance on the state assessment. Clemente is squarely in the center of the radar screen of its state accountability system, with conditions that make its engagement in improvement a very high-stakes enterprise.

Building School Capacity through State Accountability Systems

Thornton and Clemente are both improving, yet both are low-performing schools under the terms of the current accountability systems in their states. Both will almost certainly be classified as "in need of improvement" under NCLB. Moreover, there is a high probability they will, in the near term, continue to fail to meet their performance targets, although they will probably continue to improve. However, they are not failing because the people in them do not recognize their limitations or fundamentally disagree about the principles upon which the accountability system is based. In fact, people who work in both schools accept that they are not doing as well as they should.

How is it that Thornton and Clemente are improving and failing at the same time? Aren't accountability systems designed to reward and sustain improving schools, not penalize them? The reason these schools can be improving yet failing becomes evident when we look at them more closely. What we see is that the process of improvement is anything but constant and linear (see box, "The Process of School Improvement"). Both Thornton and Clemente had initial gains, but their performance has gone flat and sits below target. This is a predictable pattern through the entire improvement process when one considers what it takes to move instructional practice at scale in schools and school systems. As noted earlier, significant gains in performance are usually followed by periods of flat performance. These flat periods are important parts of the improvement process. They allow individual teachers to consolidate and deepen the knowledge and practices acquired in earlier stages; they allow schools to diagnose and identify the barriers to the next stage of improvement; and they allow time to diagnose the next set of problems and look for the capacity to work on them. However, in many accountability systems, these flat periods are seen as failures to improve and they carry heavy penalties.

"The process of improvement is anything but constant and linear."

Extreme mountain climbing provides a good analogy. Consider a group of climbers discussing strategies for an assault on a difficult 20,000-foot mountain. The goal is achievable only if each member works at full physical and mental capacity, and only if the group fully develops it capacity to work together on the assault. Would any

The Process of School Improvement

School improvement generally proceeds in phases rather than in a steady, linear fashion. Each phase involves more complex and demanding challenges than the prior phase and thus more skill and knowledge on the part of people in the organization. Although none of these phases is inevitable or occurs in every school, each captures the flavor of what improvement can look like from the vantage point of schools as they are going through a particular phase.

Improvements in school performance seem to follow these phases.

- √ **Problem Recognition**—Schools recognize and internalize problems of performance by paying attention to evidence on student performance. They choose a proximate performance target—increasing reading performance, for example—and focus their work on improving individual and organizational capacity to meet this target.

- √ **Low-Hanging Fruit**—If schools succeed in choosing the right target and developing the initial knowledge and skill in teachers and students to reach that target, they typically see a modest bounce in student performance. Often, these initial moves, especially in very low-capacity schools, consist of very low-level changes, such as devoting a set number of minutes per day to teaching reading, realigning the curriculum so content that is tested is taught before the test is given, or identifying and focusing on students whose performance could easily be improved and whose improved test results would make the entire school look better. This can be called the “low-hanging fruit” phase, or, less complimentarily, the “some teaching versus no teaching” phase.

- √ **Stagnation**—These initial simple moves usually turn out to have very short-term, very disappointing effects. Yet the critical point is the school has decided to make some collective commitment to a goal that has to do with performance. This is a prerequisite to developing internal accountability—namely, the capacity for individuals in the school to subscribe to common expectations about what it is important to work on and what students should be able to do. If the organization reads its performance well, it is during this phase that schools often try to tackle a more ambitious kind of instructional improvement. This improvement is often focused on adopting specific curricula and instructional practice.

- √ **External Help**—This next phase almost always requires the school to receive some external help, support, and professional development. New practices take time to acquire and implement with any consistency. They also require people to organize and manage themselves around increasingly clear, collective goals—another defining characteristic of internal accountability. Yet schools that go through this phase almost always see gains in student performance. This is partly because they are learning to work together more powerfully and partly because they are teaching different content in different ways. Just as predictably, performance tends to

go flat again almost immediately. This is where Thornton and Clemente are in the improvement cycle. This happens because the problems of improving student performance are more complex than the strategies adopted during this phase can cope with.

√ **Barrier Resolution**—The school most often chooses the next problem to tackle based on an analysis of the barriers to continued improvement. Typically, the problems that schools work on during this phase require some external help to resolve. They are problems like increasing the consistency and cognitive demand of instruction or determining why the instructional strategies they adopted earlier work for some students and not for others.

√ **Impossible Work**—The problems of improvement become more complex and demanding as performance increases; the challenges to existing instructional practices and existing organizational norms become more direct and difficult. Often schools experience a crisis at this point, where teachers and principals argue that the work has become impossible to do with existing resource constraints and that expectations set by external accountability systems are simply impossible to meet. What teachers and administrators are saying is they simply do not have the capacity to make the next round of changes, and they usually are correct. This is the impossible work phase. The conditions for future improvement are present in the school, but the capacity to make improvement is not. It is critical for schools to receive high levels of support during this phase—to get help in diagnosing the next problems, to get assistance from people with expertise about problems of student learning and instructional practice, to broaden and deepen common expectations about high-quality instructional practice, and to see schools in similar circumstances that have managed to move through this phase.

√ **Transformed Organization**—Schools that make it through this crisis typically emerge as much different organizations—stronger, more coherent, with responsibilities more widely distributed, with much higher morale related to student learning, and with much higher cognitive demand in the classroom. Yet they often have difficulty demonstrating that these changes are consequential because going through a crisis saps the energy and commitment of people while it is going on, and often performance goes flat during these periods. This occurs because the school has built the capacity for higher-level instruction, but it has not yet seen its full effects. This is where more concentrated work on instructional practice—not less—is important because teachers and administrators must understand that not only have they changed the way instruction occurs in the school, but they also have changed their own capacity to take responsibility for, as well as manage their school's response to pressure for, performance.

√ **Self-Management of Improvement**—The next phase of improvement is one that few schools achieve, even nominally "high-performing" schools. During this phase, the school collectively takes over managing its own improvement process, and administrators, teachers, and students internalize the values of managing and monitoring their own learning.

reasonable mountaineer suggest that all members of the team head straight up the mountain, as a group, in one continuous assault? This would be a formula for disaster. To be successful, the assault must be carefully staged to provide time for the group to pause, pool its knowledge of the mountain, regain its physical capacity, regroup, and agree on a plan for the next phase. The more effective the group, the more skillful it is at finding and using these periods of consolidation.[4]

Performance targets for many low-performing schools are unattainable using their existing capacities. Most schools, even nominally high-performing schools, couldn't do this work using their existing capacities. To meet these performance targets, schools must develop successively higher capacities. Each new set of capacities addresses the next level of problem. Each level of increased performance carries its own new set of problems. Each new level of capacity requires a period of consolidation. Acknowledging the gap between capacity and performance in accountability systems isn't an argument for abandoning performance targets altogether. It is, however, an argument for bringing capacity-building measures into better alignment with performance measures in the design of accountability systems.

Recommendations for Governors and Other Policymakers

Accountability systems should be designed to push and support schools like Thornton and Clemente to get better. To improve accountability systems, governors and other policymakers should take the following actions.

- **Develop an infrastructure at the state and local levels to provide expertise and support to low-performing schools**. Improvement is a process, not an event. Schools don't suddenly "get better" and meet their performance targets. Schools build capacity by generating internal accountability—greater agreement and coherence on expectations for teachers and students—and then by working their way through problems of instructional practice at ever-increasing levels of complexity and demand. At each successive stage, the work at the next stage can look impossible. This process must be managed by experts and informed by people who have worked with schools confronting similar problems in other settings.

  Building capacity in low-performing schools will require a lot of feet on the ground—people who know something about school improvement. Schools like Thornton and Clemente need help at a particular time, with a particular set of problems relating to a specific set of instructional issues. They need help in diagnosing the problems of instructional practice they are facing and what to do about them. They need this help, not on a one-time basis, but over time, in a sustained way, as they develop their own capacity to diagnose and solve instructional issues. This assistance requires people who know content and pedagogy, not generalists. It requires instructional resources (materials, coaching of teachers, and work with school and district support staff) that schools and districts may not have. It will require state departments of education with greater capacity to lead school improvement. Building such an infrastructure takes time and careful investment in human resources. It also requires policymakers, districts and schools to develop measures of instructional improvement and performance that are much closer to the classroom than the state assessments that are the basis of accountability systems.

- **Ensure accountability systems can distinguish among schools that are improving and those that are not.** Most state accountability systems make no distinction between low-performing schools that have demonstrated the capacity to improve instruction and performance and those that have not. Consequently, states have no way of identifying and treating differently low-performing schools that are improving. Accountability systems should not penalize schools that are making significant improvements in instructional practice and demonstrable gains in student performance. These schools are the repositories of the knowledge that the rest of us will need to understand how to deal with the schools that are not yet on an improvement path. Incorporating this idea into state policy will require finer-grained assessments of school performance and a recognition that improvements in capacity and performance do not occur in a constant, linear fashion. A school that demonstrates

a significant gain in performance during one testing period, for example, should not be required to maintain a constant gradient of performance over several successive periods. School performance should be averaged over long enough periods to allow schools like Thornton and Clemente to take credit for previous gains during periods when their performance is flat.

- **Require school districts to develop systemwide instructional improvement strategies.** The school district—not the individual school or the state—is the logical entity to assume long-term responsibility for monitoring the performance of low-performing schools and developing systems of instructional support around them. A state infrastructure for working with low-performing schools is only as good as the capacity of local school systems to use it. Many school districts with high numbers of failing schools have developed systemic strategies for instructional improvement; some have not. Federal law now requires districts identified for improvement to develop plans to address the professional development needs of their teachers and principals and to allocate not less than 10 percent of Title I funds toward professional development.[5] States should use these plans to ensure districts develop systemic strategies for instructional improvement. Among the characteristics of such strategies that states should stipulate as criteria for approval are:

  - systemwide curriculum content in reading, writing, and mathematics;
  - professional development systems that reach every teacher in a given content area; and,
  - annual school-level performance plans that set targets for professional development, improvement of practice, and performance.

  Districts receiving support from state sources should be expected to enhance their own capacity to help schools improve.

- **Create incentives to stabilize the population of teachers and administrators in low-performing schools**. Low-performing schools often operate in chaotic and unstable environments. With high turnover of personnel, investments in the knowledge and skill of teachers and administrators has a limited effect. Schools such as Clemente face a labor force in which 15 percent to 25 percent of its teachers turn over in the course of a single year. Teachers and administrators who demonstrate skill are at a premium and are recruited away to settings where they feel conditions are more conducive to their success. What Clemente and similar schools are doing is, in effect, subsidizing surrounding, higher-performing schools and districts through its investments in its own workforce.

  Help can come from stronger incentives for teachers to stay in these schools. Districts and states should partner with local systems to create and implement these incentives. Suggested incentives include:

  - *financial* (e.g., loan forgiveness for teachers in low-performing schools, extra compensation, academic improvement bonuses, housing bonuses);
  - *professional* (e.g., professional development for teachers and administrators, access to advanced degree programs); and
  - *improved working conditions* (e.g., induction programs for new teachers, reduced teaching loads, improved physical infrastructure and materials, eliminating residency requirements).

- **Continue to examine and improve the design of accountability systems.** A fundamental design problem with current accountability systems is the lack of correspondence between the expectations for annual improvement and what we know about how schools improve. This design weakness makes it more difficult for schools like Thornton and Clemente to improve. Whether or not the annual performance increments required to meet that target are reasonable, Thornton and Clemente are judged by that standard. This condition makes it extremely

Improving Accountability Systems

States are using the provisions of NCLB to develop innovative approaches to hold schools accountable and build capacity. **Mississippi**, for example, has made growth in student performance one component of its accountability system and plans for meeting performance requirements under the new federal law. **Indiana** will average the three most recent years of test score data and compare the average to the most recent year of scores. The higher of the two scores will be used for school accountability.

On the issue of Adequate Yearly Progress (AYP), several states (including **Delaware**, **Georgia** and **Wisconsin**) are using annual accountability targets that rise more rapidly in later years. Other states (including **Florida** and **South Carolina**) are spacing performance targets out so that these targets rise in three-year increments. By delaying or spacing performance targets, states allow schools more time to improve instruction, implement professional development strategies and build teacher and administrator capacity before they are required to make large performance gains.

States have taken a range of approaches to incorporating federal performance requirements in their state accountability systems. Some states have integrated AYP into their state accountability systems. **Ohio**, for example, places its schools into five categories: "excellent," "effective," "continuous improvement," "academic watch," or "academic emergency." A school that misses AYP can receive the highest label in the state accountability system—"excellent"—if the school meets Ohio's additional indicators**.** If the school misses AYP for three consecutive years, the highest rating a school can receive is "effective."

States are also measuring school performance using different indicators. **Florida**, for example, will use a return on investment indicator that links dollars spent to student achievement, and individual student progress toward standards in addition to AYP. By integrating individual student progress into school accountability systems, Florida and other states allow schools missing AYP by wide margins to demonstrate improvement. **Ohio**'s additional indicators include citizenship tests, science tests and attendance. Length of time that a school has been low-performing is also an element of Ohio's accountability system.

Sources: *Mississippi Consolidated State Application Accountability Workbook*, March 19, 2003, 8*; Indiana Consolidated State Application Accountability Workbook*, January 6, 2003, 23; *Delaware Consolidated State Application Accountability Workbook*, January 31, 2003, 32; *Georgia Consolidated State Application Accountability Workbook*, May 1, 2003, 22; *Wisconsin Consolidated State Application Accountability Workbook*, May 20, 2003, 16; *Florida Consolidated State Application Accountability Workbook*, March 26, 2003, 22; *South Carolina Consolidated State Application Accountability Workbook*, January 31, 2003, Attachment D; *Ohio Consolidated State Application Accountability Workbook*, January 6, 2003, 10.

risky and difficult for Thornton and Clemente to engage in any improvement effort that does not guarantee immediate gains, potentially eliminating many of the capacity-building measures that would ultimately turn these schools around.[6]

Many accountability systems require schools with the lowest capacity to make the largest gains. There is little or no slack in the system to adjust performance expectations to an understanding of the process of improvement or the limited capacity of schools at the low end. The solution to this problem may lie in experimenting with a variety of different ways of setting performance targets and measuring performance and scaling performance expectations to the realities of school improvement. Several states have begun this process of experimentation in the accountability plans they are required to produce under NCLB (see box, "Improving Accountability Systems").

Conclusion

Low-performing schools are part of a larger problem of educational accountability. Solving the problem will require accountability systems that can distinguish between schools that are improving and those that are not. It will require that systems move beyond compliance and sanctions to build the capacity of schools to continually improve. Finally, it will demand state intervention strategies that do more than get people to do the right thing. They must focus instead on the difficult, challenging, and important task of figuring out the right thing to do.

Endnotes

[1] Lynn Olson, “All states get federal nod on key plans,” *Education Week*, 18 June 2003

[2] The schools described in this paper are real; their identities have been changed to protect their confidentiality.

[3] Charles Abelmann, Richard Elmore, with Johanna Even, Susan Kenyon, and Joanne Marshall, *When Accountability Knocks, Will Anyone Answer?,* Consortium for Policy Research in Education (CPRE), 1999; and Martin Carnoy, Richard Elmore, et al., *The New Accountability* (N.Y.: Routledge Falmer, forthcoming).

[4] Another useful analogy is aerobic conditioning. People in modest physical shape do not work themselves into great shape in one continuous path of improvement. Substantial periods of increased cardio-vascular capacity are followed by (usually discouraging) flat periods in which it seems that no improvement is occurring at all. In fact, what we know is that during these periods the cardio-vascular system is literally opening up and enlarging new capillaries to carry oxygen-enriched blood to muscle tissue—building capacity. Then, as if by a miracle, the exerciser suddenly notices that she is able to operate at a much higher level of conditioning.

[5] U.S. Department of Education, *No Child Left Behind Act of 2001*, http://www.ed.gov/legislation/ESEA02/pg2.html#sec1116

[6] Performance expectations under NCLB are not entirely linear—states can hold performance standards for schools constant for up to three years as long as state targets rise to 100 percent proficiency by 2013-14.

National Governors Association
Center for Best Practices
444 North Capitol Street, Suite 267
Washington, D.C. 20001
202.624.5371
www.nga.org