The downs and ups of accountability in California. Will the state's political leadership stay the course this time around?

California, normally a bellwether state, its trends and culture spreading as fast as buzz on the latest Steven Soderbergh flick, was a late arrival to the modern school-accountability movement. Even the 1994 federal Title I reforms, which required states to develop the three major prongs of an effective accountability system (academic standards, tests linked to the standards, and a mixture of assistance and sanctions for low-performing schools) did little to stimulate California into action. In fact, it wasn't until the results of the 1994 National Assessment of Educational Progress (NAEP) in reading were released that the state got serious about accountability. California's absolutely disastrous performance—it tied Louisiana for last place among the 37 participating states—was a source of deep embarrassment to a state that had long prided itself on its K–12 public schools and the University of California's eight undergraduate campuses, arguably the finest system of public universities in the world.

Analysts have cited a legion of reasons for the state's slide in achievement: the steady leaching of resources from the schools that was the inevitable result of the infamous 1970s property-tax revolt led by Howard Jarvis; a long period of economic woes caused by layoffs in the defense industry; curriculum experiments with "whole language" reading instruction and "new math" that were at best a distraction and at worst quite damaging; a school finance lawsuit that led to a dramatic increase in the state's authority over school budgets and operations; and a massive influx of new students and non-English-speaking immigrants that almost surely depressed test scores. Whatever the reasons,
the result was a sharp drop in the public's and policymakers' confidence in the abilities and, indeed, the motivation of local educators. A political coalition and consensus developed around the ideas that accurate information on student and school performance is needed in order to hold educators accountable and that educators can't be trusted to work hard without the existence of positive and negative state incentives.

Since then, California has developed the basic foundations of a coherent accountability system. Nevertheless, despite the progress, many aspects have yet to fall into place at the school level, and there is always the danger that the underlying political coalition will collapse and take the accountability system down with it. Education reform generally, and in California especially, seems to follow a pattern of taking one step up and two steps back, as successes in leadership and shifts in the political winds lead to the dismantling of earlier reforms and the layering on of new ones. This is certainly true of the accountability movement in California, which has a tortured history of reform and retrenchment that for years left the state with no real measure of how its students were performing. So the question is, will California stay the accountability course this time? Will the system survive the almost inevitably low passage rates that will occur when the state shifts to a new high-school exit test this year? Will the state not only punish low-performing schools, but also give them the resources and technical assistance necessary to build solid academic programs? And will the new provisions of the federal Elementary and Secondary Education Act (ESEA) hamper or support Californian's efforts to improve schools?

Experiments Gone Awry

California may be a latecomer to the modern accountability movement, but the state has been no stranger to the idea that standardized tests ought to be used to gauge school performance. California state officials began to lose confidence in local school authorities and teachers as early as the 1960s. This concern culminated in the California Assessment Program (CAP), which was first administered in 1972. The program required each student to take a sample of questions from an overall test, a method called "matrix sampling." Matrix sampling did not produce scores for individual students, but it did permit an in-depth assessment of a school's performance in each subject. Each school was given a grade based on a comparison between its actual scores and its predicted scores, which in turn were based on its students' socioeconomic backgrounds. Any newspapers published the schools' grades.

In the 1980s CAP became the victim of a political dispute between Republican governor George Deukmejian and the elected state superintendent of education, Bill Honig, a Democrat. Honig signaled that he was considering running for governor against Deukmejian, which led to the governor's vetoing of CAP. Local educators did not mourn the death of CAP, since many local schools had received negative publicity as a result of their low test scores. They also contended that CAP did not assess what teachers were teaching. The effectiveness of CAP relied on the willingness of parents, educators, and school boards to use information as a tool for school improvement. Policymakers especially hoped that angry parents would exert pressure on low-performing schools and school boards, but their hopes were never realized.

It wasn't until a decade later that Californians again had detailed information on the state's academic performance. In the early 1990s, Honig created the California Learning Assessment System (CLAS). While CAP had focused exclusively on multiple-choice tests, CLAS asked students to read a poem or passage and respond to questions like: "Pick a part that is especially interesting and explain your reasons," or "What are your feelings about this poem?" Educators, teachers especially, liked the test's open-ended and creative nature, but rumors quickly spread among some parents that the tests contained "objectionable content" that threatened moral values and students' privacy. Other critics questioned whether open-ended items could be objectively scored and thus serve as a reliable measure of school performance. What bothered Republican governor Pete Wilson, however, was that CLAS didn't provide scores for individual students. Wilson believed that individual test scores enhanced parental oversight and responsibility for their children. In the end, despite the support of many educators, CLAS could not overcome its technical and political problems—state superintendent Honig, the major supporter of CLAS, was convicted of a felony and removed from office—and it succumbed to Governor Wilson's veto in 1995.

Californians were again plunged into darkness, as the state returned to allowing each school district to choose its own test. So it came as quite a surprise when the state's vaunted school system finished lower than Mississippi's on the 1994 National Assessment of Educational Progress (NAEP). State policymakers, led by Governor Wilson, reacted with alarm and passed a new state standards and assessment pro-
In 1996, the state implemented the Standardized Testing and Reporting system, or STAR. New procedures were designed to give the governor and other public officials more influence over test questions, addressing concerns that had plagued the California Longitudinal Assessment System (CLAS). A commission was established to develop "academically rigorous" standards in all major subject areas at every grade level. The majority of the commission's members were appointed by the governor. A six-person statewide review panel was to review all test items to ensure they were free of questions about:

- A student's or parent's personal beliefs, sex life, family life, morality, or religion;
- Personal characteristics such as honesty, integrity, sociability, or self-esteem.

One result is that the new California Standards Tests (known as the CST), which were first administered in 2001 but won't enter the accountability system until 2002, rely only on closed-end multiple-choice questions, unlike the CLAS test.

The 1996 law envisioned a logical sequence of events. First, the state would create and approve curriculum standards in each grade, and then a set of assessments would be developed that were linked to the state's curriculum. However, California's policymakers, concerned about the lack of performance information and accountability in the system, decided that the state couldn't wait that long. So, starting in 1997, all students in grades 2 through 11 were required to take the Stanford 9, a commercially available, nationally normed test. This satisfied Governor Wilson's desire to promote parental responsibility by sending home test scores for each child as soon as possible, but the Stanford 9 rapidly became the tail that wagged the accountability dog. As the accountability system continued to unfold under the next governor, Gray Davis, monetary incentives for teachers and schools were attached to test-score gains on the Stanford 9. Schools got the message and began preparing students for the material tested by the Stanford 9—even though they were supposed to be teaching the state's curriculum in order to prepare for the California-developed assessments that were to come later. Thus as usual, and not by design, the tests drove the standards and the curriculum rather than the other way around.

Foundations

Davis, a Democrat, ran for governor in 1998 with education as his top priority, proclaiming, "Local control of California education is a disaster." His theory of change relied on rewards and sanctions for schools along with an informed public agitating for improvement. After Davis's election, he delivered on his promise to centralize accountability at the state level by persuading the legislature to pass legislation incorporating four basic elements:

- An academic performance index to measure each school based on its changes from year to year;
- An intervention program for underperforming schools;
- Monetary awards for schools making gains and college scholarships for high-scoring students; and
- An exit exam that all students must pass to graduate from high school.

So far, the only component of the academic performance index has been the Stanford 9, since these were the only data available (see Figure 1). Each California school receives a ranking on the index from 1 to 10, and schools are also ranked against the schools in their income range. The release of the rankings

Mixed Results (Figure 1)

Younger students have made the biggest gains under California's STAR accountability system. The reading scores of older students have actually slipped.

![Graph showing Stanford 9 Math and Reading Test results for grades 5 to 11 in 2001](source: California Department of Education's Education Data Partnership)
is a major media event and the rankings are published on the Internet. The state’s standards-based assessment will be incorporated into the index this year. Other data, such as attendance and dropout rates, will hopefully be reliable enough to be added in the future.

The intervention program is reserved for schools that both fall below the Stanford 9 national average and do not meet their goal for gains on the Stanford 9. Test scores are disaggregated by race and ethnicity, but schools need only to raise their overall scores to avoid the intervention program. The state sets a goal of about a 5 percent gain for every school in the state. The growth target is set even higher if a school is grossly underperforming on the Stanford 9. A part of the intervention program, grants of $50,000 per school were first awarded to 304 schools during the 2000-01 school year. The funds were used to hire consultants who help plan and evaluate school improvement, followed by a $200 per-pupil implementation grant, to be used for anything from motivational programs for teachers to new instructional programs. The consultants are often retired teachers and school administrators, as well as staff from the U.S. Department of Education’s regional education laboratories. Early evaluations of these plans have not been encouraging: many of the plans use strategies aimed at eliciting short-term test-score gains rather than inspiring long-term school improvement.

Schools that have failed to meet their growth target after two years of implementation are deemed low performing. This can trigger the state’s intervention in various ways, including having the state superintendent take over a school. There are criteria for triggering a state intervention, but it is less clear how a school can get itself out of the doghouse. When is a school no longer judged to be low performing? There is also continuing disagreement over which schools should receive extra resources. Governor Gray Davis wants to spread the school-improvement grants among all schools that are below the average—roughly 50 percent of the schools in the state—because he wants to build a centrist political coalition, but his fellow Democrats want to concentrate funds on the 20th percentile and below. The outcome will depend on the state’s budget, however. State interventions may have to be postponed in a recession because of the need to cut funding, but the legislature is more willing to cut the rewards than the funds for state interventions.

In 2001 Governor Davis also provided $667 million in school-performance awards for test-score gains on the Stanford 9. In order to receive awards, schools must produce both overall gains and gains for various ethnic and racial groups. One program gives each teacher in a school $25,000 if the school greatly exceeds the state target for Stanford 9 increases. An additional $135 million was provided for college scholarships to 11th graders who scored high on the Stanford 9, despite concerns that the Stanford 9 tests mainly basic skills, while college-bound high schoolers are expected to take courses focusing on English literature and algebra.

The final element of the current accountability system is that by 2004 all students must pass a high-school exit exam that is based on the state standards for grades 7 through 10. This is more of a minimum-skills test than an exit exam, but it at least includes two essays.

Wilson and the Democrat-controlled legislature believed California schools were in such bad shape that they could not wait for the sensible sequential process of developing a distinctive California test based on California’s curriculum standards. This was the equivalent of saying, “Fire, ready, aim.” The controversy has not died out, even though scores on the Stanford 9 have been going up every year since 1998, and advocates are claiming victory. For example, Ron Unz, who sponsored a successful state initiative to restrict bilingual education, issued a press release saying that English immersion was a major cause of the rising scores. However, the state legislature’s independent Legislative Analyst’s office, headed by Elizabeth Hill, warned that increases in scores could be the result of the fact that questions on the Stanford 9 remained similar for three years, so students and teachers are becoming more familiar with the style of questions.

By 2000, the entire state leadership realized that something had to be done to better align all the facets of the accountability system and to lessen the impact of Stanford 9 testing. The California-developed, standards-based assessments were first administered in 2001. These include secondary-school end-of-course exams that, like New York State’s Regents exams, are arousing the interest of colleges and universities as they look for ways, besides the SAT and grades, to measure students’ ability and motivation. Substantial funds have been appropriated for professional development and textbooks aligned with the standards-based assessments. Standards for teacher training were recently aligned with the state’s curriculum and assessments. For now, a shorter form of the Stanford 9 will be used in most subjects, and it will have less impact on the incentives and sanctions that are being doled out. In three years, California may become a national leader of the accountability movement, as all the major components of a coherent and rational accountability system fall into place.

**More Obstacles**

Of course there is still much work to be done. The policymaking is largely over, but the campaign to change classroom instruction has just begun. Local educators are keenly aware of the state’s accountability pressure, but awareness does not equal commitment to the state’s goals or to classroom change. The accountability system must win the support of teachers, and schools and
teachers need the training and resources necessary to help them teach the state's challenging standards. Unfortunately, California is running out of money just as its accountability system is coming together. The energy crisis drained the state of its surplus that was generated by the technology boom of 1995–2000, and the economy, dependent as it is on the fortunes of Silicon Valley, isn't what it was three years ago. The state surplus had helped to expand standards-based professional development programs, including algebra academies for teachers and students. But Governor Davis had to cut the teacher institutes in 2001, and the state is forecasting a $12.5 billion deficit in 2002.

The governor proposed to cut the 2002 program of incentive payments to teachers and schools by 60 percent, and the legislature voted to cut it by even more. Meanwhile, differences across schools in per-pupil funding continue to create differences in learning opportunities, and this issue will only become more visible as the full accountability system is implemented and poor schools fall in large numbers. The state is again facing a lawsuit, this time brought by the American Civil Liberties Union, claiming that it is not providing an adequate education to all children.

There is also the challenge of moving past the strong signals that have been sent to educators that the Stanford 9 test is the accountability system. Moreover, local educators claim that there are still too many secondary-school tests, including four statewide assessments (the Stanford 9, the high-school exit exam, the standards-based exams, and the Golden State Exam, a holdover from the 1980s). California's colleges and universities add two admissions tests (SAT I and II); and the University of California, California State University, and the community colleges use three different placement exams. These assessments somehow need to be rationalized.

Policymakers must also deal with the inevitable backlash. Statewide, only 1 percent of students have opted out of testing, but a very vocal set of parents in San Diego and Marin County have refused to let their children take the state test. These represent only a fraction of the parents in high-income suburbs who believe that standards-based accountability will only lead to more state interference, which will push the teaching in their school to the lowest common statewide level. Wayne Johnson, president of the California State Teachers Association, has called the high-school exit exam a "disaster." He predicts that students whose English is limited will fail in large numbers and claims that the math questions are too difficult. A few teachers rejected their Stanford 9 incentive bonuses, calling them "blood money" and "bribery." Earlier concessions to the teacher unions forbadethe use of state tests for hiring, firing, or promoting teachers. Students' grades are based primarily on teacher-designed or local tests, so the stakes for students are unclear.

Adding to the confusion is the new federal education law that President Bush signed in 2002. The major problem is a federal requirement that by the fall of 2002 all teachers hired under Title I must be "highly qualified," and by 2005 every public school teacher must be "highly qualified." In 2001, 42,427 teachers (one out of seven statewide) were working in California without a preliminary credential that the state has defined as a minimum requirement. Several California districts hire teachers who have just a bachelor's degree and a passing score on a minimum skills test that is set at the 10th grade level. California officials have no strategy to fulfill this federal teacher quality requirement.

While the spirit of California's accountability system is in accordance with the federal law, a significant issue is the definition of pupil "proficiency." In 2001, California established five performance levels on its state test and defined "proficiency" as the level of achievement necessary to enter a university. Only 30 percent of California students are at the proficient level, but the federal law requires all students to be "proficient" in 12 years. California must either lower its standards for proficiency or negotiate a federal waiver. Moreover, California's accountability law measures annual pupil growth on a school-wide basis, averaging all student scores in a school, while the federal law requires each student to meet growth targets.

California's experience with school report cards also raises questions about the federal proposals. The 2002 federal law requires that states provide annual report cards with a wide range of information for each district and school. California uses both an elaborate report card (with 50 criteria) and a school ranking (on a scale of 1 to 10) based solely on test scores. The long report card has the virtue of being comprehensive, but it is not used by educators and is too complex for the public. The ranking, by contrast, is easily grasped by parents and is noticed. A single number, however, oversimplifies the complexity of schooling. Parents should receive a few clear measures and an indication of where to get more information about their child's school. Overall, report cards don't cause much change, and the more data they contain, the more confusing they are to parents.

In the end, perhaps the instability of earlier accountability schemes will help to keep this one in place. Policymakers keep saying that California must stick with something, and the business community has formed a coalition to preserve the current agenda and direction. The real political test will come when students begin to fail the graduation exam, or when schools face dramatic interventions like faculty firings or reorganizations. The key issue is whether the state's accountability system will change and improve classroom instruction. So far, the linkage between state policy and the classroom has been uncertain and largely indirect. Any state policymakers realize this and hope that a coordinated system will send clear signals to local educators. But it is unclear whether the current blend of exhortation, sanctions, incentives, and publicity will affect classroom instruction in a major way.

-Michael W. Kirst is a professor of education at Stanford University and director of Policy Analysis for California Education, a joint venture of the Stanford and University of California-Berkeley schools of education.