The Rise, Fall, and Rise Of State Assessment In California

1993-96

BY MICHAEL W. KIRST AND CHRISTOPHER MAZZEO

Messrs. Kirst and Mazzeo offer their postmortem on the short-lived and controversial California Learning Assessment System.

QUESTIONS about the feasibility of and political support for new forms of pupil assessment have become major issues. With the California Learning Assessment System (CLAS), California became a pioneer in these new forms of assessment. For a variety of reasons, however, parents, conservative religious groups, the California School Boards Association, the California Teachers Association (CTA), and the governor all raised objections to the assessment during its 1993 implementation.

As a result of this dissent, CLAS is now discontinued, but many questions still remain. Answers to them can shed light both on the future of assessment policy in California and, more generally, on the politics of testing. What happened to CLAS? Why did it generate so much opposition? Why was CLAS not able to sustain the political coalition that created it? What are the future prospects for testing policy?

The CLAS case illustrates some of the difficulties involved in large-scale transformation of state assessment systems. For advocates of performance-based testing, the case stands as an exemplar of the difficulties of moving policy toward more "authentic" forms of assessment and away from

MICHAEL W. KIRST is co-director of Policy Analysis for California Education (PACE) and a professor of education at Stanford University, Stanford, Calif., where CHRISTOPHER MAZZEO is a graduate student. *They wish to thank the Consortium for Policy Research in Education and PACE for supporting this research and Neal Finkelstein for his assistance in preparing the manuscript. This article was originally presented at the annual meeting of the American Educational Research Association, New York, 1996.*

Illustration by Jonathan Bouw

the measurement of basic skills through multiple-choice exams. While factors unique to California (e.g., election-year politics) can partially explain the fate of CLAS, other aspects of the case offer more general lessons for reformers about the politics of testing policy in the United States.

CLAS was developed in 1991 to replace its predecessor, the California Assessment Program (CAP). CLAS was designed to satisfy a number of needs that the previous testing program had not met. Three goals of CLAS stand out: 1) to align California's testing system to the content of what was taught in schools, as represented in state curricular frameworks; 2) to better measure attainment of curricular content through performance-based standards and assessment; and 3) to provide assessment of individual student performance as well as data on schools and districts. The test was intended to create comparable scores for all parts of the state's education system. The performance of these discrete parts of the education system would be measured through both on-demand assessments given once a year and portfolios that would keep track of student work over a longer period of time.

The Rise and Fall of CLAS

Controversy over CLAS intensified after the first round of tests was given in the spring of 1993. Rumors quickly spread among conservative groups and parents about the test's "objectionable content." These rumors were exacerbated by the secrecy that shrouded the assessment — secrecy that the California Department of Education said was essential for retaining the integrity of the items and avoiding the expense of developing new ones.

Without actual exams available, the rumors gathered momentum, and with them came complaints by religious groups that the test's content undermined parents' moral values and invaded the privacy of students and their families. While some parents complained about privacy, others took issue with both the open-ended nature of the performance assessments and the lack of "objective" scores. (See the sidebar on page 323 for a sample test.) The designers of the CLAS items had not included any potential critics of the test — that is, those who represented traditional religious and conservative groups.

The first official response to the CLAS controversy came in January 1994, when State Sen. Gary Hart put together S.B. 1273, the CLAS reauthorization bill. The new bill took four steps to deflect the criticism of the tests. First, a review panel would be appointed to ensure compliance with the intent of the legislation. Second, past copies of the test would be provided each year for review by the public. Third, school boards would be allowed to review each year's test before it was given — provided that they guaranteed test confidentiality. Finally, to answer concerns about open-ended assessments, the bill increased the number of fact-based, multiple-choice, and short-answer questions. Though the bill was an honest attempt to deal with the controversy, it eventually contributed to the demise of CLAS later in 1994.

A series of events in subsequent months in 1994 also played a major part in the end of CLAS. The state board of education's removal of a selection by Alice Walker from the 1994 reading test sparked a firestorm of negative reaction in newspaper editorials and from such groups as People for the American Way. Then the scores of the 1993 tests were released in March. Some schools that had done well on previous assessments had fared poorly on the new tests. And some of these schools were in the wealthiest areas of the state. The results increased anger on all sides. In April, the *Los Angeles Times* published an investigation critical of the sampling procedures. The article claimed that there were more than 11,000 sampling violations in 1993. School boards in Conejo Valley and Antelope Valley in Southern California opted out of the 1994 tests. And the conservative Rutherford Foundation filed suits on behalf of parents in Sacramento and San Bernardino, claiming that the tests violated privacy laws.

The final heated blow in April came in a scathing letter from Del Weber, president of the CTA, to William Dawson, acting superintendent of public instruction for the state. Weber's letter rebuked the state department for both the design and the administration of the assessments. While ultimately supportive of the CLAS concept, CTA's reaction added to the public relations nightmare for CLAS.

At the end of the month, the state department responded. In a press release dated April 30, Dawson addressed the criticism of the previous months. Stating in strong language that all districts would be required to administer the tests, he noted that they could create opt-out procedures for parents who wished to do so. Defending both the confidentiality of the assessment and the scoring procedures used in the first year, Dawson claimed that the *Los Angeles Times* article had been inaccurate. Only 150 schools had samples that should not have been released to the public. Nonetheless, Dawson vowed to have the public more involved in future test review. Most important, he commissioned a scholarly review board of testing experts, led by Lee Cronbach of Stanford University, to examine sampling and other statistical issues from the 1993 tests.

In early May the governor finally spoke out. Emphasizing the controversy over content and the sampling problem, Gov. Wilson called for the state auditor general to review CLAS fiscal issues. Maureen DiMarco, secretary of child development and education, called the assessment "seriously flawed" and "disastrous." Responses to the comments by Gov. Wilson and Secretary DiMarco were swift. In a May 12 article in the *Los Angeles Times*, Bill Honig, the former state superintendent, blasted Wilson and his aide for jumping off the CLAS bandwagon. Implying that the governor did so for political gain, Honig claimed that Wilson's actions played into the hands of extremists. In the ensuing months the verbal volleys flew. In mid-July, the state department put the 1993 test items on public view. Initial reports were positive, as many parents who had expressed fears claimed that the tests were not as bad as they had originally believed. But any boost the state department might have received from the public viewing was soon nullified by the release of the expert statistical review committee's report.

While Dawson and his department tried to put a positive spin on it, the report of Cronbach's group, the Committee on Sampling and Statistical Procedures, was undeniably critical. Suggesting that operational problems were significant in 1993, the committee recommended some measures to ensure technical competence and quality control in future tests. While the samples were basically sound, the committee found them poorly implemented. Regarding school site scores for 1993, the assessment was found to have inadequate reliability, and concerns about large standard errors led to the recommendation that future school-level assessments be admin-

Conflict over new performance-based assessments is not unique to California.

istered on an experimental basis.

In his press release announcing the report, Dawson emphasized the positive and implied that the department's plan regarding both technical procedures and individual scores had been validated by the committee. But Gov. Wilson and DiMarco did not see it that way. Citing some of the conclusions of the expert report, the governor vetoed S.B. 1273 on 27 September 1994 and called for a new statewide testing program to replace CLAS. Wilson's veto announcement showed him moving away from his earlier emphasis on the problems of sampling and content and focusing on the fact that the bill made no provisions for individual scores for students. Individual scores could enhance parental responsibility and school advising.

In her comments, DiMarco claimed that the new bill veered away from the intent of the original CLAS bill — S.B. 662 — which made individual pupil scores the overriding goal of CLAS. Instead, the department had given priority to the performance-based aspects of the test, and this decision was codified into the new bill.

In a sense, the comments of the governor and DiMarco were correct. An analysis of the two different CLAS bills reveals many instances in which references to individual scores have been removed or changed. Indeed, the part of S.B. 662 that DiMarco cited regarding the primacy of individual scores was deleted from the later bill. The ambiguity comes from a reading of the initial language of that section, which states that "comparable individual pupil results shall be completed prior to any expansion and development, or both, of new performance-based assessments except to the extent that performance-based assessments are an integral part of the system for providing individual pupil results."

It can be argued that, in the state department's judgment, performance-based assessment was an "integral" part of providing individual pupil results and would therefore take priority. Certainly that is what Honig believed when CLAS was created in 1991. Wilson and others, however, saw the priorities differently. Given the political controversy, it is not surprising that their view won, even though CLAS was supported by most major education groups in the state, including the CTA.

Why CLAS Was Discontinued

Gov. Wilson's veto was merely the final blow to a new testing system that had had difficulties from the beginning. Certainly, such political factors as the strength of traditional religious groups and the governor's election-year need to shore up his support with them helped to undermine CLAS. But the CLAS case also highlights a number of more general issues regarding the politics of assessment policy in the U.S. Conflict over new performance-based assessments is not unique to California; Virginia, Arizona, and Connecticut have recently faced similar controversies. The demise of CLAS offers a constructive lesson for policy makers who are committed to assessment reform rooted in performance-based testing. Three key dimensions of the CLAS case stand out as lessons for testing policy in general: 1) the tension between political and technical factors, 2) the divergent priorities and goals of key stakeholders, and 3) the extent of antigovernment feelings among the public.

Tension between political and technical factors. While there is much agreement among policy makers and testing experts on the benefits of performance-based testing, policy makers subordinate technical realities to political ones. In the CLAS case, the political reality dictated an overly optimistic time line for implementation in 1994 — against the recommendation of those familiar with performance-based examinations. The traditional needs for a wide-scale assessment — test validity and reliability — are more problematic given the state of the art of performance-based and constructed-response exams. Developing an assessment that measures the complex skills detailed in curricular frameworks is a difficult and costly process. Making such an assessment a high-stakes affair for students and schools raises the ante considerably on technical and cost issues. As the statistical review committee noted in its report, the tradeoff between cost and precision in a performance-based exam is significant. Making scores reliable and valid for accountability purposes is a difficult proposition.

Furthermore, the committee noted that a design that performs well in assessing schools creates difficulties for measuring individuals. The chances that students will get comparable forms of the test decrease with a larger sample, which makes student-level accountability decisions risky and possibly quite unfair. Yet the state department was expected to solve these problems and deliver a test with student and school scores by 1993. The department's decision to push performance-based testing at the expense of individual scores says much about its priorities. Still, it is likely that whatever choice the state department made would have alienated some groups. Policy makers' need for quick and decisive action may prove to be disastrous for performance-based reforms such as CLAS.

Divergent priorities and goals of key stakeholders. Like all policies, assessment policies are created by political coalitions. Since the actors involved often have divergent goals for testing, it is often necessary to write legislation in vague terms or to incorporate seemingly conflicting goals into the same policy. The three key policy makers who helped to create CLAS — Wilson, Hart, and Honig — all had very different priorities for the testing program.[1] Gov. Wilson's top priority was to replace the older CAP system with a new one that provided individual student scores, allowing for more parent awareness and stringent accountability for teachers. State Sen. Hart was much more interested in holding the schools accountable for performance. In keeping with the ideas of the National Governors' Association and other policy organizations, Hart wanted to trade the schools' deregulation for stricter performance accountability. Finally, Honig, then state superintendent, and the state education establishment were committed to performance-based testing and to tying assessment to the curricular frameworks.

All these goals appear in the initial legislation. However, once implementation began, it was clear that not all the priorities

could be accommodated. When the state department implemented a policy closest to the vision of Honig and Hart, the governor and others who supported his position balked. The controversy over testing content helped strengthen the opponents' contention that the test was "seriously flawed." What has not been resolved — in California or in other states — is the matter of the goals of assessment policy. Should tests emphasize student- or school-level accountability? Are assessments predominantly informative and persuasive tools to help students and teachers perform better? Or are they regulatory instruments tied to rewards for good scores and sanctions for bad ones? The California case failed to answer these questions, and this failure led to an inevitable conflict once CLAS was implemented.

Antigovernment feelings. Many policy makers have been surprised by the extent and virulence of the negative reaction to such reforms as performance-based assessment and outcomes-based education. Since many of the loudest protests have come from religious groups, they have often been dismissed as mere "extremism." However, this ignores the origins of much of the unrest: antigovernment feeling. Nearly all the criticism of CLAS was directed at the staff of the state department or at other key figures in the state capital. Much of the criticism focused on the privacy issue. As one of the lawyers for a parent group that sued the state put it:

> The state has an interest in assessing the quality of teaching in the schools. [It] also [has] an interest in knowing whether kids can think rather than regurgitate facts. But there's a difference between testing a student's ability to think and asking [students] what they think about personal things. And frankly, the latter is no business of the state.[2]

The criticism did not stop at privacy concerns. In an editorial the *Orange County Daily News* railed against the "Sacramento bureaucrats" to whom CLAS cedes control over "core issues of schooling."[3] The president of one of the school boards that opted out of CLAS claimed that the concern was "not the moral issue as much as the absence of testing basic skills." These criticisms illustrate the extent of the antigovernment feelings in California at the very time that reforms are trying to expand the reach of the state and persuade people of the need to rethink traditional ways of testing. The convergence of these two trends does not bode well for ambitious testing reform in other states.

The Rise of California Assessment

CLAS ended with several issues unresolved and a well-organized opposition that also objected to California's participation in the 1994 federal Goals 2000 program. Consequently, it is surprising that a new state assessment passed in a single year, especially since the Republicans had gained enough seats to obtain a one-vote majority in the California Assembly. Each of the major roadblocks, however, was overcome by a winning coalition of Democrats in the Assembly and in the Senate, the governor, the education groups, and big business.

The three major issues from the CLAS debacle remained, but new approaches were fashioned. A two-track assessment system was designed to develop individual-level assessment and state-level (school- and district-level) assessment. The student track consists of districts using commercially available tests, with the state providing $5 to each district for each student who takes the test. Presumably, these commercial achievement tests will be aligned with state curricular frameworks and textbooks that are being revised in 1996.

School- and district-level assessment is to be phased in by 1998-99 through a new instrument that is to include a balance of "basic and applied skills." The bill's authors expect that the assessment of applied skills will include performance assessments, while the assessment of basic skills will emphasize multiple-choice responses. In order to satisfy political demands for more emphasis on the basics, grade-level curriculum standards are to be formulated as a guide for teaching and performance levels.

The state will explicitly agree on how good is good enough for subject attainment in each grade. It is not clear how the new state assessment at the state and local level will be aligned with individual student tests from commercial publishers. The expectation is that both assessment tracks (state and pupil) can be linked to the state's curricular frameworks as well as to the National Assessment of Educational Progress (NAEP).

These substantive changes in assessment could pass the legislature if public confidence in state government can be rebuilt after CLAS. The governor and conservative parent groups wanted more control over the design and implementation of state assessments. The state department of education squandered its legitimacy during the CLAS dispute and needed to be less prominent. Indeed, Delaine Eastin, the newly elected state superintendent, pledged to the legislature in public hearings that the state department staff that designed CLAS would not be involved in any new state assessment.

Two new governance mechanisms for assessment were designed to provide more gubernatorial and citizen influence and to lessen the state department's visibility. A 21-person Commission for the Establishment of Academic Content and Performance Standards (hereafter, the Commission) will be responsible for developing "academically rigorous" standards at every grade level in all major subject areas. The majority of the Commission's members will be appointed by the governor.

The public will also participate in the approval of the tests themselves. A six-person Statewide Pupil Assessment Review Panel will review all tests to ensure that they contain:

- no questions about a student's or parent's personal beliefs about sex, family life, morality, or religion; and
- no questions designed to evaluate such personal characteristics as honesty, integrity, sociability, or self-esteem.

A majority of this Review Panel must be parents with children in public schools. Legislators and local school board members can review the contents of any approved or adopted test, as long as they agree to maintain the confidentiality of test items. Easily understood materials describing the nature and purpose of the tests must be made available to members of the public, including parents and students.

The new California assessment was approved by a crucial Assembly committee by a single vote. The governor seems satisfied and has provided adequate funding in his recommended 1996 budget. While some of the assessment provisions could be inconsistent, there has been rising concern about the attainment of pupils since California tied with Louisiana for last (out of 37 states) in fourth-grade reading on the 1995 NAEP. The new state assessment must be comparable to the NAEP and must

Sample Reading Test

Read the poem that appears below and answer (on another sheet) the questions that follow. Feel free to make notes in the margins as you read.

Introductory Note: This poem, inspired by the Statue of Liberty, was written in 1883 by Emma Lazarus. Lazarus was born in Eastern Europe and immigrated to America as a young woman. Like most immigrants of that time she came to the United States by boat, entering the country through the port of New York, where the Statue of Liberty stands in New York harbor (between New York City and Jersey City, New Jersey) at the entrance to N.Y. and the USA. The poem became famous when it was later inscribed on the base of the Statue of Liberty. The original "Colossus," one of the seven wonders of the ancient world, was a huge statue that straddled the harbor of Rhodes in ancient Greece.

The New Colossus (1883) Notes

Not like the brazen giant of Greek fame,
With conquering limbs astride from land to land;
Here at our sea-washed, sunset gates shall stand
A mighty woman with a torch, whose flame
Is the imprisoned lightning, and her name
Mother of Exiles. From her beacon-hand
Glows world-wide welcome; her mild eyes command
The air-bridged harbor that twin cities frame.
"Keep, ancient lands, your storied pomp!" cries she
With silent lips. "Give me your tired, your poor,
Your huddled masses yearning to breathe free,
The wretched refuse of your teeming shore.
Send these, the homeless, tempest-tossed to me,
I lift my lamp beside the golden door!"

1. What is your initial reaction or response (your thoughts, feelings, observations, questions, ideas, etc.) to this poem?

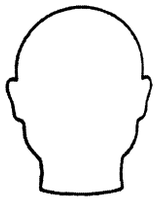
2. Pick a line in this poem that seems to you especially important or interesting. Write out the line and then explain your reasons for selecting it.

3. How do you interpret the name "Mother of Exiles" in line 6? What is the significance of this name in the poem?

4. The last part of the poem says, "Give me your tired, your poor."

a) Who is "me" and who is you or "your?" Who is speaking in the last 5 lines of this poem and to whom are these lines addressed?

b) Using the "Open Mind" outline provided, show with drawings, symbols, or words what the speaker of these lines is thinking or feeling or what a person hearing these lines might be expected to think and feel.



c) Explain your graphic.

5. Use the opportunity provided by this question to say anything else you might want to say about this poem. You might want to talk about its form or language, its meaning to you personally or as a member of a group, its cultural or historical or ideological or aesthetic significance, or anything else you haven't already said about the poem.

Source: Sheridan Blau, Department of English, University of California, Santa Barbara. Reprinted with permission.

include an appropriate balance of types of assessment strategies, including multiple-choice items, short-answer questions, and applied writing. The state assessment will report on grades 4, 5, 8, and 10, but performance standards must be established for every grade level.

Opponents will focus on holding up funds for developing the new state assessment, but they appear ready to let the individual testing proceed in 1996. The final political battle stemming from CLAS is far from over. Soon after the bill was signed by the governor, Orange County conservatives attacked it. "I say kill it," said Joan Wonsley, a Dana Point mother of three and co-founder of an anti-CLAS parent group. "They want to know what kids think. They're getting psychological, talking about political correctness. They're reshaping social attitudes."[4]

State assessments involve questions that are high-stakes politics: What knowledge is most worth knowing?[5] Institutionalizing new forms of assessment will need to draw on a reservoir of public trust and public understanding, but creating this reservoir will require more than state-level political marketing and campaigning. California's experience suggests that an elite professional alliance cannot both set the agenda for reform and persuade the public that this agenda is best. However, merely responding to what the public desires does not capitalize on assessment research and the growing knowledge base of professional educators.[6] Somehow, education leaders must find a middle ground that bridges grassroots opinion and improved assessment concepts. This will require more than engagement or interactions between the public and professional educators. Developing new assessments must entail guidance and leadership, combined with a grasp of how the public interprets the various messages that it hears about testing.

1. Lorraine McDonnell, "Assessment Policy as Persuasion and Regulation," *American Journal of Education*, vol. 102, 1994, pp. 394-420.
2. *Ed Cal*, weekly newsletter of the Association of California Administrators, 30 May 1994, p. 8.
3. Ibid.
4. John Gittelsohn, "All Aren't Hailing CLAS Replacement," *Orange County Reporter*, 15 October 1995, p. 3.
5. See Michael W. Kirst, "The Politics of Nationalizing Curricular Content," *American Journal of Education*, vol. 102, 1994, pp. 383-93.
6. Robert Rothman, *Measuring Up* (San Francisco: Jossey-Bass, 1995).