TESTING

SNAPSHOTS OF EIGHT STATES

Over the past 10 years, forces seeking reform in education have worked to require increased state and/or local testing. In many places, this movement followed widespread dissatisfaction with the quality of education as personified by perceived ability levels of graduates. In response, public and community leaders began to seek "accountability" from schools — specific statements of what is being attempted and specific measure ments of what is being accomplished. Often, the Governor or the state legislature became a critical player in this movement. Concerned over the need for a well-educated work force in the national competition for jobs and industry, states have increasingly turned to testing.

Educators, often initially alarmed by demands for increased testing, have in most instances moved from opposition to cooperation, and have worked to design tests and test environments conducive to learning. Two forms of testing have increased; these are minimum competency testing and assessment testing.

Minimum competency testing seeks to determine whether or not students are learning the information defined in that system as basic. Minimum competency testing normally comes in tandem with opportunities for help to those failing the tests and opportunities for re-testing. In time, pass rates for minimum competency tests rise substantially over initial levels.

Assessment testing is quite different, in that it seeks to measure the effectiveness of various school programs. Assessment testing is more informative to educators and cheaper than the traditional standardized tests. Using specific modern quantitative techniques, assessment testing can be accomplished using a relatively small number of students. Thus, money is saved in test instruments and processing, and substantial time is saved by leaving most students in class. Assessment testing is generally thought to be

a useful comparison between programs in different schools, because it is designed to measure program or school effectiveness, not simply the comparative ability levels of students.

In order to accurately convey the various forces behind the current testing movement, OTA asked individuals in eight states to describe, in their own words, the recent history of testing in their state. The following papers are presented unedited, and are intended to give a flavor of the many ideas and circumstances at work in different states, and the various approaches that states have adopted.

A BRIEF HISTORY OF STATE TESTING POLICIES IN CALIFORNIA

Susan M. Bennett and Dale Carlson
California Assessment Program
California State Department of Education
January 1986

Prepared Under Contract With
The Office of Technology Assessment
Congress of the United States

A BRIEF HISTORY OF STATE TESTING POLICIES IN CALIFORNIA

Origins of State Testing: 1961-1964

Statewide achievement testing in California originated in 1961 with the recommendation of a citizens' advisory commission. The commission recommended that the Legislature set a level of instruction through the State Board and the "mandatory statewide examinations be utilized to establish this standard" (Joint Interim Committee, 1961, p. 38). The assessment program first implemented in 1962 embodies the concept mandated in 1961 and implemented for the first time in 1962 embodied the concept of accountability, but did not set standards in a literal or uniform sense More than a million students — the entire student population at grades 5, 8, and 11 — were tested annually from 1962-1964 in reading, language, mathematics, and intelligence ("scholastic aptitude"). Districts selected standardized instruments from lists of state-approved tests for each grade level

1965-1973

The establishment of a statewide reading improvement program in 1965 (Miller-Unruh Basic Reading Act) was accompanied by substantial modifications in the scope of content assessed and in the grade levels tested. The new legislation required districts to administer a uniform test to all students in grades 1, 2, and 3 to provide data for selecting those districts most in need of reading specialists. The legislation also instructed the State Board to adopt uniform tests at the upper grade levels; to change the grade levels tested from 5, 8, and 11 to 6 and 10; and to restrict achievement testing to a single content area: reading. An explicit proscription on public release of test results included in the 1961 testing law was reversed in 1968 when new legislation mandated that results be reported annually on a district-wide basis. Further

modification of the law in 1969 (California School Testing Act) changed the upper grade level to be tested from 10 to 12 and expanded the content tested to include basic skills in language and mathematics as well as reading. During this period districts purchased, administered, and scored the standardized test adopted for each grade level by the State Board. They returned the results to the State Department of Education to be summarized and reported to schools, districts, and to the State Board.

1973-1978

Widespread dissatisfaction with the statewide testing program — especially the resentment among district personnel of what they perceived as unfair comparisons based on commercially-produced tests that were poorly matched to the skills taught in California — led to a complete restructuring of the testing program. New law in 1973 incorporated detailed recommendations of a legislative advisory committee on testing chaired by Lee Cronbach. Foremost among the committee's recommendations was the separation of local and statewide testing into distinct programs, with the statewide program mandated to provide data for evaluating instructional programs at the school, district, and state levels, but not to provide data for individual students or classes. Multiple-matrix sampling was recommended to provide reliable data on a broad array of curricular objectives while reducing the time required for testing from three or four hours to approximately 35 minutes.

The new state-level testing program, the California Assessment Program (CAP), was first fully implemented in 1974-75 with all testing costs absorbed by the state. The design, development, and procedures of the new program were unique in the nation. CAP tests were developed for grades 1, 2, 3, 6, and 12 with the full participation of statewide committees of content area experts and classroom teachers. Each test was designed to assess specific objectives representing the full breadth of content that should be taught in each content area at the appropriate grade level. The newly-developed tests included

a grade 1 entry level test of prereading skills (to replace the end-of-year reading achievement test), a single test of reading achievement to be administered in grades 2 and 3; and tests of reading, mathematics spelling, and language for grades 6 and 12. Following the multiple-matrix design recommended by the legislative advisory committee, large numbers of items were distributed over 10-18 nonoverlapping forms for three of the new tests: the grade 2 and 3 reading test and the surveys of basic skills for grades 6 and 12. Each student at these grade levels completed a single form of the appropriate test and the results were then aggregated to provide a wide variety of program diagnostic scores for each content area and for subskills within each content areas. Scores were aggregated and reported at the school, district, and statewide levels.

The new approach to statewide achievement testing, with its focus on the assessment of school-level programs rather than the needs or progress of individual students, relegated testing for other purposes to a variety of district-level testing programs. Thus, local districts assumed full responsibility for standardized achievement testing to satisfy program evaluation requirements, to compare local performance with national norms, and to report student-, class-, and school-level scores to parents and local school boards. Legislation in 1976 and 1977 also made districts responsible for conducting proficiency (minimum competency) testing in reading, writing and computation and for developing or selecting appropriate tests to do so. Performance indicators and examples of minimum standards for testing once between grades 7-9 and twice between grades 10-11 were set by the State Board, with minimal course requirements for graduation prescribed by law. Individual districts set their own graduation standards. (Further legislation in 1981 mandated that summer school be required for all students in grades 7 to 12 who failed to meet their district's standards.) District-conducted proficiency testing was also required once between grades 4-6 to identify students in need of remediation.

Legislation in 1975 also mandated an early exit" proficiency test, the California High School Proficiency Examination (CHSPE). The CHSPE is an optional, four-hour examination that provides the opportunity for students who are 16 years old or second-term sophomores to verify their competency in basic reading, writing, and arithmetic skills. Candidates with passing scores are awarded a Certificate of Proficiency that is equivalent by State law to a high schooi diploma. Although the State Department of Education is officially responsible for the development and content of the CHSPE, it is administered by a private testing service. The CHSPE is related to CAP, the statewide testing program, only peripherally — normative data on the CAP twelfth-grade test are used as a partial basis for setting and monitoring the passing score (Carison, 1979).

1979-1982

A number of changes to CAP recommended by the 1977 Assembly Advisory Committee on Statewide Testing became law in 1978. The most significant of the changes ended testing in grade 2 and shifted resources to grade 3 to measure skills in written language, mathematics, and reading, with a heavy emphasis on comprehension. The new Survey of Basic Skills: Grade 3 was developed by staff of the State Department of Education with extensive involvement by advisory committees of content area specialists and by teachers throughout the state. First administered in 1979-80, the new test consists of more than 1,000 items operationalizing objectives found in the statewide curriculum frameworks, state-adopted textbooks, and skill areas commonly taught in California schools. Following a multiple-matrix design, items in each content area were assigned to 30 unique forms, each comprised of 34 items and requiring no more than 35 minutes for a student to complete.

A scaled score system based on item response theory was introduced for reporting the results of the new Grade 3 Survey. The new system permitted year-to-year comparisons independent of statewide performance or item changes and also permitted

direct comparisons of performance across content areas without translation into normative scores. Beginning in 1980, grade 3 school reports have included scale scores for each of the three content areas and 90 specific skill areas presented in a program diagnostic format that encourages the use of information on relative strengths and weaknesses for modifying local instructional programs.

CAP staff begin developing a new, more demanding Survey of Basic Skills: Grade 6 in 1980 following the same procedures as were followed in constructing the grade 3 test. The new Grade 6 Survey was administered for the first time in 1981-82. Each student completes one of 40 unique matrix forms consisting of 31 questions in 30-35 minutes. The new grade 6 school reports, like the grade 3 reports, provide program diagnostic information indicating relative strengths and weaknesses as shown by scale scores for the three content areas of reading, written expression, and mathematics, as well as for numerous subskills within each content area.

1983-1986

California's new Superintendent of Public instruction, Bill Honig, was elected in November, 1982, on a reform platform calling for a return to a traditional academic curriculum and to instructional practices — including rigorous testing — that represent "what we know works in education" (Honig, 1985, p. 6.). Excellence in education, as envisioned by Mr. Honig, involves preparing all students — both college- and noncollege-bound — to compete successfully for jobs that require brains rather than brawn, and eIevating them intellectually and morally through exposure to a common, irreducibie core of knowledge in the arts and sciences.

To initiate the long-term process of reform required to operationalize this vision of quality education, the Department of Education requested additional funding from the legislature and proposed a number of statutory changes. The educational reform measure passed by the California legislature and signed by Governor Deukmejian in 1983 provided

$850 of the $950 million dollars in the Department of Education's original request along with a package of 65 reforms (Hughes-Hart Educational Reform Act), including mandated graduation requirements for all students, a longer school day and year, money for textbooks and summer school, tighter discipline and dismissal procedures, and definition of statewide curriculum standards. To provide for systemwide quality control, the reform measure mandated modification of the existing statewide assessment program to emphasize higher-order academic skills and to assess additional grade levels and content areas. It also established a new end-of-course examination program to measure and reward high-level achievement in critical high school courses.

The changes in statewide testing by Hughes-Hart in 1983 reflect a general policy that standardized tests aligned with statewide curriculum objectives should be used to the greatest possible advantage to achieve the goals of curriculum reform with students of all types. More specific policy goals clarify several separate, but related, ways in which standardized tests are expected to promote curriculum reform.

1) *Standardized tests are expected to focus the attention of educators" and policy makers at all levels on the knowledge, skills, concepts, and processes which are essential for success in the more demanding high-tech job market of the future, for responsible citizenship, and for personal fulfillment. The core of content and skills to be spotlighted represents a rigorous curriculum in the humanities, natural sciences, and math and emphasizes higher-order skills such as those required to analyze complex relationships, draw inferences, and reason deductively. Although it is assumed that in practice, the scope and pace of the curriculum will reflect differences in aptitude and intelligence (Honig, p.202), it is also assumed that the majority of students are not working up to their potential, and that it is the responsibility of the schools to challenge them to do so — both for their own good and for the good of the society.*

2) *Scores on standardized tests (along with indices of performance such as enrollment in selected academic courses, the amount of homework completed on a nightly basis, and the frequency of writing assignments) provide baselines against which schools are encouraged to set targets for improvement and to complete with themselves and with other schools serving similar populations, thus tcheting the whole system upward over time toward the goal of academic excellence" (Honig, 1985, p.124).*

3) *By helping to clarify a sense of common purpose, by focusing attention on the challenging academic objectives of the reform movement, by raising expectations, and by providing feedback on improvements in achievement, standardized tests are expected to contribute — along with the curriculum they represent, more interesting and challenging textbooks, and other key components of the reform package — to rekindling a sense of excitement and enthusiasm for learning in teachers and student alike.*

4) *Standardized testing is expected to provide measures of accountability that are essential to gaining and maintaining cooperation and support for the educational reform movement from parents, educators, policy makers, the business community, and other important segments of the public. Evidence of continuing i reprove ments in student performance is expected to sustain enthusiasm over the anticipated 5-10 year period needed to fully implement the goals of curriculum reform.*

Unlike the testing reforms that have been instituted in other states in the past several years, the revisions, expansions, and additions to California's statewide testing program do not include an emphasis on minimum competency testing. On the contrary, the recent changes in statewide testing indicate a commitment to go beyond narrowly-focused tests of basic skills or minimum competencies to instruments that will truly embody the objectives of a challenging academic curriculum, measuring the full range of higher-order academic skills and using testing approaches other than the traditional multiple-choice format wherever possible.

Consistent with the legislative mandate, statewide testing has been expanded to focus instruction on the most important objectives of the reform movement and to provide accountability to the public for a more rigorous instructional program. One major component of the expansion involves additions to the California Assessment Program. CAP has added to its survey series since 1983 by developing the Survey of

Academic Skills: Grade 8, first administered in 1983-84. A matrixed test of 36 70-item forms, the grade 8 test consists of reading questions based on passages from literature, science, and social science emphasizing higher-level comprehension; questions on written expression based on student essays related to the reading passages; mathematics questions assessing computational abilities, problem solving, prealgebra, and pregeometry skills; history-social science questions emphasizing critical thinking skills as well as content knowledge; and science questions requiring knowledge of process as well as content. Tests of history-social science and science will also be developed to supplement the existing CAP surveys of reading, written expression, and mathematics at grade 6 and other grade levels as the legislature makes funds available. Other anticipated additions to the statewide testing program include a Grade 10 Surveey with grade-appropriate content paralleling that of the new grade 8 test (not yet funded by the legislature), and a direct (essay) assessment of writing skills, now in its second year of development and scheduled to be added to the Grade 8 Survey in 1987 and to the grade 12 and grade 6 tests in subsequent years.

Current efforts to upgrade the California Assessment Program's survey series also focus on the development of a completely new, expanded, and more demanding grade 12 test to replace the instrument that has been in use since 1974. The new Survey of Academic Skills: Grade 12 will be a multiple-matrix test with content in reading, written expression, mathematics, history-social science, and science. The items will assess important higher-level thinking skills and competencies identified in each of these subject areas by the Model Curriculum Standards: Grades Nine through Twelve adopted by the State Board of Education in 1985. The new grade 12 test is scheduled for partial implementation (three content areas) in 1987-88 and full implementation (including tests of history -social science, science, and a written essay) in 1988-89. The CAP surveys for grades 3, 6, and 8 will be reviewed for consistency with statewide curriculum objectives and revised as needed after the Model Curriculum Guides for kindergarten through grade 8 are completed in 1986-87.

The Golden State Examination Program (GSEP) is a second major component of the plan for expanding statewide testing to focus instruction on the curriculum objectives of the educational reform movement. Golden State Exams will be developed to measure achievement in 17 academic subjects under statewide standards of competency and to identify students qualifying for a special honor designation on their high school diplomas. Students will be tested on a voluntary basis upon completion of courses in mathematics, laboratory sciences, United States history, English literature and composition, foreign languages, and health sciences. The first two GSEP exams in beginning algebra and geometry will be field tested in 1985-86 and fully implemented in 1986-87. GSEP exams" in United States history and biology are now in the initial stages of development, The full series of tests will be developed and operationalized as funds are available.

A third component of the plan for modifying statewide testing to better meet California's educational objectives involves development of a comprehensive assessment system that will provide student-level scores to meet proficiency requirements and specialized local needs as well as provide the school-, district-, and state-level results needed for program evaluation by CAP. The proposed system would consolidate CAP'S statewide testing program with district testing programs in order to reduce the overall costs of testing, reduce the amount of instructional time devoted to testing, and ensure that testing is focused on the priorities of California's curriculum. Preliminary work has been completed, but full development of the system will require further legislative initiative.

Use and Impact of Statewide Testing

The statewide testing program, as required by the legislation that established CAP in its present form in 1973, provides group-level information to school districts, to the legislature, and to the public to be used in each of three major ways: 1) to evaluate the

effectiveness of school programs, 2) to allocate resources to schools with the greatest educational needs, and 3) to identify successful practices. This is done annually through a series of reports including school-level and composite district-level reports, a four-year school and district sum mary, and an annual report of statewide results.

In practice, CAP data are used by school personnel, the legislature and State Department of Education staff, and the public in a great variety of ways. The following are examples of some of the most common uses by each of these audiences:

1) *Educators in districts and schools typically use CAP data to evaluate strengths and weaknesses in particular content and skill areas, at specific grade levels, in particular subgroups of students, and in particular schools. Trends across years, trends across grades, and comparisons with statewide performance and with the performance of other schools serving similar students populations are also frequently emphasized.*

   Results of a survey of more than 4,600 elementary principals in 1979 indicate that most of them were using CAP results to examine curricula more closely, to develop instructional strategies to correct problem areas, to call attention to problem areas not previously noted, and/or to develop or focus teacher in-service activities. The changes principals most frequently related to CAP results include modifications in the goals and objectives of instructional programs, articulation of curriculum and teaching activities within and across grade levels, modifications in the amount of time devoted to teaching various skills, and development of new instructional materials (California, 1980). Local educators also frequently use CAP data to document the need for special funds or for participation in special projects. Recent comments by local and district administrators, both in the press and in conversation with CAP staff, indicate that they continue to use CAP data in all of the ways documented by the 1979 survey.

2) *Legislators and State Department of Education staff typically use CAP data to evaluate instructional programs and practices by examining yearly achievement in major content areas and by making comparisons of trends across content areas, across grades, across years, and across subgroups of students (classified by gender, mobility level, English language fluency, socioeconomic level, and ethnicity, as well as by supplementary information on reading outside of school, homework assignments, writing assignments, TV exposure, etc). Statewide results are also compared with national performance based on studies equating CAP tests to various nationally standardized tests as well as to NAEP.*

Statewide CAP scores indicating curriculum weaknesses have prompted intervention at the state level. For example, the relative weakness in computational skills apparent in statewide CAP results in the late '70s led to revisions of state Curriculum Frameworks and to the adoption of new, more balanced textbooks. More recently, a decline in eighth grade CAP scores in 1985 (as well as the students' below-average standing relative to national norms) has led to the formation of a Middle Grade Task Force composed of students, parents, educators, and representatives of business and industry. The Task Force, formed in January, 1986, will hold hearings throughout the state to address issues including students' maturation patterns, teacher credentialing, grade level configurations, and effective teaching strategies in order to develop a plan for improving the quality of middle grade education in California.

3) *Legislators and staff of the State Department of Education also typically use CAP data to evaluate the impact of special state and federal programs, to document need and allocate resources, to study funding models and effective schools, and to identify promising practices. Recent examples include: CAP scores in reading and mathematics (1979-1984) used as indicators of program effectiveness in comparing elementary and secondary school participants and nonparticipants in the School Improvement Program (California, 1985); CAP achievement scores used to identify exemplary schools (California, 1977; Fetler Carlson,1985); CAP twelth grade data ' used to identify low-performing high schools and their characteristics as a basis for proposing further legislation to assist such schools (California, 1984); and year-to-year improvements in CAP twelfth grade scores used to determine cash rewards to schools under the Education Improvement Incentive Program begun in 1984.*

4) *Since CAP data at the school-, district-, and statewide levels and comparisons of state results with national norms are widely reported in the press, they are major contributors to the general public's evaluation of California's schools. Parents typically use such data to make comparisons between schools and districts and realtors typically use them to argue the merits of investment in areas with high assessment scores (Powell, 1981).*

Consistent with the policies of California's educational reform movement and the accountability plan instituted in early 1984, standardized test data have been given greater influence in the past several years. In addition to the detailed information on achievement scores in CAP'S annual school, district, and statewide reports, CAP scores in reading and mathematics are now also reported at all levels of the school performance report first issued by the Department of Education in fall 1984. The high school performance report includes CAP scores as well as information on students' SAT scores, College Board Advanced Placement examination scores of 3 or above, and College Board

achievement test scores on selected examinations. These test data along with other statewide performance indicators are now being used to recommend California schools for the Federal School Recognition Program. They will also serve as the primary basis for selecting schools for the new California School Recognition Program, the next phase of the accountability program to be implemented.

California's Education Improvement Incentive Program (EIIP) has also increased the emphasis on standardized test data in the past several years by offering a cash incentive for improvement achievement on the CAP twelfth grade test. Enacted as a part of the Hughes-Hart educational reform bill in 1983, EIIP is not a part of the Department of Education's accountability program. Nonetheless, by distributing awards of over $14.6 million to more than half of the high schools in California, EHP has focused a great deal of attention on statewide testing at grade 12. New legislation has recently been introduced to extend the incentive program to the sixth grade level.

Summary

It would be premature to attempt to assess the impact of the changes in statewide testing mandated by California's 1983 educational reform legislation at this point. Major test development efforts are underway on the new grade 12 test, direct assessment of writing skills, and the Golden State end-of-course examination program (see above), but the first of these new assessment instruments will not be implemented until 1986-87, and the full set of Golden State Examinations may not be finalized for a number of years. Parts of the grade 8 test — the first of the new tests to be completed — have been in place since 1984, but the science component will be added for the first time in spring 1986. In California, as in the other states that are now beginning to implement educational reform, the appropriate time to look for improvements in achievement attributable to expanded testing programs and to the variety of other reform measures instituted concomitantly is still a year or two down the road (Kirst, 1985).

In the meantime, California's state testing program is contributing to the goals of the educational reform movement by focusing attention on statewide curriculum objectives, by providing a basis for schools to set targets and better their performance from year to year, and by providing accountability to the public. The California Assessment Program is, by design, well suited to perform these roles and has been doing so for a number of years by reporting broad and comprehensive program diagnostic information to educators at all levels, to the legislature, and to the public. Publicity surrounding the educational reform movement in general, the new statewide curriculum standards, the accountability program with its performance reports, the new tests being developed, and the Educational Improvement Incentive Program, have all heightened awareness of the existing testing program. Evidence provided by newspaper reports throughout the state, orders for rationale and content documents" for the CAP tests, and attendance at workshops held to introduce the new grade 8 tests and to assist teachers in using program diagnostic data to evaluate strengths and weaknesses in their instructional programs indicate that educators are seriously concerned about their performance on the CAP tests. One consequence of this concern is that districts are taking steps to incorporate higher-level thinking skills and other competencies identified by the statewide curriculum standards in their local programs.

References

California Legislature. (1984). Overcoming the Odds: Making High Schools Work. Sacramento: Assembly Office of Research.

California State Department of Education. (1977). California school effectiveness study. the first year: 1974-75. Sacramento: State Department of Education.

California State Department of Education. (1980). Student achievement in California schools: 1979-80 annual report. Sacramento: California Assessment Program, California State Department of Education.

California State Department of Education. (1985). Report of Consolidated Application Programs 1984-85. Sacramento: Program Evacuation and Research Division, California State Department of Education.

Carlson, D.C. (1979). Statewide assessment in California. Studies in educational evaluation 5, (pp. 55-75). Great Britain: Pergamon Press.

Fetler, M.E. & Carlson, D.C. (1985). Identification of exemplary schools on a large scale. In Research on Exemplary Schools (pp. 83-96). Academic Press.

Honig, B. (1985). Last chance for our children: How you can help save our schools. Reading Mass: Addison-Wesley.

Kirst, M.W. (1985). Sustaining state education reform momentum: The linkage between assessment and financial support. (Policy Paper No. 85-C3). Standard, CA: Institute for Research on Educational Finance and Governance.

Joint Interim Committee. (1961). Report of the Joint Interim Committee on the public education system. Sacramento: Senate of the State of California.

Powell, M. (1981)., Uses of state assessment information. In D.C. Carlson (Ed.), Testing in the states: Beyond Accountability (pp. 13-29). San Francisco: Jossey-Bass.