The Limits of State Mandated History Curriculum: California’s Standards and Tests for History-Social Science

Bradley Fogo

Abstract

This paper examines the California Standards Tests for History-Social Science. It begins by tracing the development of California’s standards-based reforms for history-social science over the past twenty years – from the adoption of the History-Social Science Framework in 1988 to the latest developments in state testing policy. The paper then describes the format and organization of the California Standards Tests, their technical development and how they are used by the California Department of Education. Finally, the paper explores issues of validity and the alignment of test items with the history standards they measure and argues that the California Standards Tests suffer from questionable validity and may promote troubling ways of teaching and learning history.

This past spring, 8th, 10th, and 11th graders in California public schools took the California Standards Tests (CSTs) for History-Social Science. Implemented initially in 2001, these exams consist entirely of multiple-choice questions aligned with the grade-level content and analysis standards of the History-Social Science Framework for California Public Schools. The purpose of these criterion-referenced tests, according to the California Department of Education, is “to measure how well students...are achieving state-adopted content standards.” Performance standards divide test results into five categories: advanced, proficient, basic, below basic, and far below basic. The Department of Education’s goal is to have all students in the state achieve the level of “proficient” on these tests.

Despite calls for independent evaluations of state testing systems, analyses of the California history tests are missing from the literature. Recent research on state history assessments has largely overlooked California. This oversight may be attributed to the relatively low stakes of the state’s history tests. Results of these tests do not figure into student promotion or graduation and account for only 7 to 15 percent of a school’s Academic Performance Index (API) – the state’s measure of a school’s achievement for the year. The history-social science tests are often overshadowed by California’s “higher” stakes exams -- the California High School Exit Exam and the California Standards Tests for math and English/language arts -- which account for most of a school’s API and the Adequate Yearly Progress (AYP) provision of the No Child Left Behind legislation.


4 For a recent collection of this research, see S.G. Grant, ed., Measuring History: Cases of High Stakes Testing Across the States (Charlotte, NC: Information Age Publishing, 2006).
Nonetheless, the lack of attention paid to the *California Standards Tests for History-Social Science* is surprising. California has taken a leading role in standards-based reforms for history education over the past two decades. The *History-Social Science Framework* of 1988 was the first state mandated curriculum of the standards era to place history at the center of the social sciences. Furthermore, many organizations uphold the *History-Social Science Content Standards for California Public Schools*, adopted by the California State Board of Education in 1998 and integrated into the Framework in 2001, as exemplary. The Department of Education developed these standards to organize teaching resources, guide instruction, and help students gain “core knowledge in history” and “develop the critical thinking skills that historians and social scientists use to study the past.” Research across states and subjects, however, has indicated that state tests also influence classroom organization, instruction, and learning. Nevertheless, while California’s history standards have inspired several analyses and commentaries, few have taken notice of the California history tests.

What follows is an examination of the *California Standards Tests for History-Social Science*. In order to understand these tests, one must place them contextually within a larger system of accountability. To do this, I begin by briefly tracing the development of California’s standards-based reforms for history-social science over the past twenty years – from the initial adoption of the *History-Social Science Framework* to the latest developments in state testing policy. I then focus on the *California Standards Tests*. I describe the format and organization of the exams, their technical development and how they are used by the California Department of Education. Finally, I explore issues of validity and the alignment of test items with the history standards they measure.

Overall, this paper claims that the development and implementation of the *California Standards Tests for History-Social Science* exemplify several challenges in creating large scale, state history assessments – namely, how to create tests that are at once feasible and efficient, yet also reliable and valid. This paper concludes that the California’s history tests are both promising and problematic. On one hand, they are part of an accountability system attempting to insure a rigorous and engaging history education for all California students. However, these tests also suffer the shortcomings of relying exclusively upon multiple-choice tests to measure historical knowledge and may promote troubling ways of teaching and learning history.

The implementation of the *California Standards Tests for History-Social Science* marked the culmination of a fifteen-year movement toward standards-based history instruction and accountability in California that, in many ways, began with the adoption of the *History-Social Science Framework for California Public Schools* in 1987. The *Framework* developed from two primary reform initiatives. First, it was the product of California’s move towards “systemic reform” – a process through which the State Department of Education took a greater role in creating and aligning curriculum, assessments, and instructional materials. Legislated by 1983’s *Hughes-Hart Act* and guided initially by California State Superintendent of Public Instruction Bill Honig, systemic reform rested upon the assumption that innovative and intellectually ambitious curriculum developed at the state level would help to improve teaching and learning throughout

---


8 California State Department of Education, *History-Social Science Framework* (Sacramento, CA: California State Department of Education, 1988). In California, the term social-science is often used synonymously with social-studies.
California’s public schools. The California State Board of Education adopted the 1988 Framework to serve as the state’s curricular center for history-social science education.

Furthermore, the Framework marked an early victory for the history reform movement that emerged in the wake of *A Nation at Risk*. Written primarily by Dianne Ravitch and Charlotte Crabtree, the Framework realized many of the objectives prescribed in Ravitch and Finn’s *What Do Our 17 Year Olds Know?* and the Bradley Commission’s *Building a History Curriculum: Guidelines for Teaching History in the Schools*. It mandated the teaching of history at every grade level in California schools and devised three history curricula – primary (K-3), middle (4-8), and secondary (9-12) - with American history studied in grades 5, 8, and 11 and world history in grades 6, 7, and 10. The Framework consisted mainly of “course descriptions” for each grade level. These narrative summaries suggested the different themes and units that students should learn during the year. Although not beyond reproach, the Framework’s focus on the teaching and learning of chronological history received accolades from teachers, academics, and reformers. It quickly became, as Linda Symcox described it, “a model and rallying point for history-centered curricula nationwide.”

Indeed, in the years following the Framework’s adoption, history emerged as one of the core academic subjects for standards-based reforms both nationally and in several states. In 1992, Congress allocated funding for the creation of national history standards and, in 1994, President Bill Clinton signed the *Goals 2000: Educate America Act* that, among other things, provided funding for the development of state standards and assessments including, specifically, history as opposed to social studies. The Clinton administration strengthened this legislation later that year by requiring states to develop subject-specific standards and assessments in order to receive Title I funds through the reauthorization of the *Elementary and Secondary School Act*.

At the time of the Framework’s adoption, the California Department of Education used the California Assessment Program (CAP) to measure schools’ academic performance. Initiated in the 1970s, CAP included written and multiple-choice tests administered across several grade levels. The tests focused on reading, writing, and math and were designed, in large part, to gauge minimum competency of basic skills. The Department of Education planned to develop CAP during the 1980s to include additional content areas and “emphasize higher order academic skills.”

---


12 Similar to the arguments that erupted over the ill-fated National History Standards, opposition to the Framework focused almost exclusively on the document’s conceptions of history. However, whereas the National History Standards were primarily attacked from the Right and dismissed as overly critical of American institutions and excessively multicultural, the Framework’s most vocal critics came from the Left and charged that the state history curriculum overemphasized progress, unity, and the Christian values of American history while downplaying diversity and overlooking racism and the costs of American expansion. For examples of such criticism, see Catherine Cornbleth and Dexter Waugh, *The Great Speckled Bird: Multicultural Politics and Education Policymaking* (New York, NY: St. Martins Press, 1995); Ronald Evans, “Diane Ravitch and The Revival of History: A Critique,” *The Social Studies* 80, no. 2 (1989): 85-92, and Christine Sleeter and Jamy Stillman, “Standardizing Knowledge in a Multi-cultural Society,” *Curriculum Reader* 35, no. 1 (2005): 27-46.


1984, the Department implemented an 8\textsuperscript{th} grade test that contained science and history-social science components. Plans to create history-social science tests in grades 6, 10, and 12 were cut short in 1990, however, when Republican Governor George Deukmejian vetoed the re-authorization of CAP over a funding dispute with Democrats in the state legislature.\textsuperscript{15}

In 1991, Senate Bill 662 created the California Learning Assessment Program (CLAS) to replace CAP. Unlike CAP, CLAS tests were aligned with state frameworks. Furthermore, the Department of Education promised that CLAS would improve upon CAP by providing not only school level scores, but also test results for individual students. Designed by the progressive leaning Assessment Policy Committee, CLAS featured authentic, performance assessments aimed at eliciting higher order thinking while asking students to demonstrate understanding.\textsuperscript{16} CLAS initially comprised exams in reading, writing, and math administered to 4\textsuperscript{th}, 8\textsuperscript{th}, and 10\textsuperscript{th} graders. The Department of Education planned, as it had similarly with CAP, to develop history-social science and science CLAS exams across grade levels.\textsuperscript{17} Once again, however, these plans never materialized. Republican Governor Pete Wilson vetoed the reauthorization of CLAS in September 1994 after only two years of implementation. Wilson's veto resulted, in large part, from the spirited opposition of conservative groups critical of the language arts assessment's multi-cultural content and focus on students' personal beliefs and feelings in addition to charges that the CLAS tests lacked sufficient reliability.\textsuperscript{18}

The following year, Wilson shifted California's commitment to systemic reform by signing into law the \textit{California Assessment of Academic Achievement Act}. This legislation, in line with federal calls for standards-based reforms, mandated the creation of "academically rigorous content standards" for core subjects and the adoption of "tests that yield valid, reliable estimates of school performance and…pupil performance."\textsuperscript{19} Emerging from the uproar over CLAS and the state's dismal performance on the 1994 National Assessment of Education Progress (NAEP) tests in reading and math, the \textit{Assessment of Academic Achievement Act} contributed to what Michael Kirst and Lisa Carlos described as "an era of retrenchment," when Republicans took control of the California state assembly and sought to develop a "back to basics" approach for state mandated curriculum.\textsuperscript{20} Instead of aligning innovative curriculum with authentic assessment, the California's Department of Education re-focused its attention on developing a reliable and valid accountability system.

By 1998, the state's Academic Standards Commission had developed specific content and analysis standards for history that aligned with the \textit{History-Social Science Framework}. The Department of Education adopted the \textit{History-Social Science Content Standards for California Public Schools} on October 9, 1998 and integrated these standards into the Framework in 2001.

The addition of content and analysis standards fundamentally changed the \textit{History-Social Science Framework}. It made the document less suggestive and more prescriptive. Instead of providing general "course descriptions" to guide instruction, the Framework now included hundreds of specific items - content standards - to be taught at each grade level: 134 for grades K-5, 175 for grades 6-8, and 203 for grades 10-12.\textsuperscript{21} Each content standard contained specific historical events, people, places, and phenomena. Take, for example, the following 8\textsuperscript{th} grade and 11\textsuperscript{th} grade standards: Content Standard 8.9.5:

\begin{verbatim}

\end{verbatim}

\textsuperscript{15} McDonnell, 2004.


\textsuperscript{19} \textit{California Assessment of Academic Achievement Act}. A.B. 265, Chapter 975, February 6, 1995.

\textsuperscript{20} Michael Kirst and Lisa Carlos, \textit{California Curriculum Policy in the 1990s: "We Don't Have to be in Front to Lead"}(San Francisco, CA: WestEd, 1997), 2.

\textsuperscript{21} There are 48 content standards for 6th grade, 61 for 7th grade, 69 for 8th grade, 48 for 10th, and 70 for 11th grade.
Analyze the significance of the State’s Rights Doctrine, the Missouri Compromise (1820), the Wilmot Proviso (1846), the Compromise of 1850, the Kansas-Nebraska Act (1854), the *Dred Scott v. Sandford* decision (1857), and the Lincoln-Douglas debates (1858).22

Content Standard 11.9.3:
Trace the origins and geopolitical consequences (foreign and domestic) of the Cold War and containment policy, including the following:

- The era of McCarthyism, instances of domestic Communism (e.g., Alger Hiss) and blacklist
- The Truman Doctrine
- The Berlin Blockade
- The Korean War
- The Bay of Pigs invasion and the Cuban Missile Crisis
- Atomic testing in the American West, the “mutual assured destruction” doctrine, and disarmament policies
- The Vietnam War
- Latin American policy23

The Standards Commission also created analysis standards to help students develop “intellectual, reasoning, reflection, and research skills.”24 They divided these standards into three categories - “Chronological and Spatial Thinking,” “Historical Research, Evidence, and Point of View” and “Historical Interpretation”- and split them into three levels: twelve analysis standards for grades K-3, twelve for grades 6-8, and fourteen for grades 10-12. Examples of the high school analysis standards included: “Students distinguish valid arguments from fallacious arguments in historical interpretations”; and, “Students interpret past events and issues within the context in which an event unfolded rather than solely in terms of present-day norms and values.”25

In the spring of 1998, after three years without any statewide assessments, the Department of Education initiated the STAR (Standardized Testing and Reporting) Program. Authorized by Senate Bill 376 in October of 1997, the STAR program relied upon the national, norm-referenced Stanford 9 tests to assess the history-social science skills of 9th, 10th, and 11th graders as part of its broader mandate to test reading, language, spelling, and math for all public school students in grades 2 through 11. In 1999, the Public Schools Accountability Act formalized California’s accountability system by creating the Academic Performance Index—the scaled score of a school’s aggregate test results ranging from a low of 200 to a high of 1000. Signed into law by Democratic Governor Gray Davis, the Public Schools Accountability Act made the STAR program “high stakes” by authorizing the publication of school test results and state rewards and sanctions for underperforming schools. Today, the Academic Performance Index helps determine if a school makes Adequate Yearly Progress (AYP) under the No Child Left Behind legislation.26

The implementation of the Stanford 9 tests marked the largest increase ever for state history testing in California. However, these “off the shelf” tests were not aligned with the state’s content standards. In the spring of 2001, therefore, the Department of Education introduced the California Standards Tests for History-Social Science. Developed at first by Harcourt Education Measurement, these tests, similar to the Stanford 9, were initially administered to 9th, 10th, and 11th graders. The Department of Education awarded the testing contract for the *California Standards Tests* to the Education Testing Service (ETS) in 2002 and later replaced the 9th grade test that covered the California history standards from


23 Ibid., 52.

24 Ibid., 21.

25 Ibid., 40.

26 If a school fails to make AYP for consecutive years, it must design a program improvement plan. If it continues to fall short of its AYP, a school is subject to a number of state interventions. These may include reconstitution or state takeover of the school.
grades 6 to 8 with an 8th grade test. This year, because some schools teach the 10th grade standards in 9th grade, the Department of Education changed the 10th grade test to an “end of course” exam, allowing students to take the test upon completing the high school world history curriculum.

The California Standards Tests are criterion-referenced exams intended to measure student knowledge of the California history standards. The 8th grade exam consists of 75 multiple choice questions aligned with the standards for grades 6-8 while the 10th grade “end of course” and 11th grade tests each include 60 multiple choice questions to assess the 10th and 11th grade standards respectively. The tests are offered in English only and consist of four-option, multiple-choice items. ETS converts raw scores to scaled scores ranging from 150 to 600. For example, a raw score of 41 on the 2007 11th grade test equated to a scaled score of 378. ETS establishes cut scores that divide scaled scores by proficiency levels. For example, cut scores and proficiency levels for the 2007 11th grade test were as follows: 150-269 = far below basic; 270-299 = below basic; 300-349 = basic; 350-400 = proficient; 401-600 advanced. The raw score equivalents for these ranges were 0-17, 18-24, 25-35, 36-45, and 46-60, respectively. In order to score at the proficient level on any of these tests, a student must answer approximately 60 percent of the questions correctly.

Over the past six years, student performance on the California history tests has improved steadily. In 2002, 24% of the state’s 10th graders and 31% of its 11th graders tested at proficient or above; in 2007, the percentage of students testing at these levels rose to 29% and 35% respectively. In 2003, 27% of California’s 8th graders achieved levels of proficiency; this number rose to 35% in 2007. The average raw score for 8th graders on the 2007 test was 38.18, while the averages for 10th and 11th graders were 29.59 and 30.39.

History-social science test results do not figure directly into the determination of a school’s Adequate Yearly Progress. Moreover, history-social science is not included in the California High School Exit Exam, which focuses on math and English-language arts. The history tests do account for approximately 7 percent of middle schools’, and 14 percent of high schools’, Academic Performance Index. Despite such relatively low stakes, the Department of Education assumes these tests will serve as a diagnostic on how students are achieving standards and therefore help teachers adjust instruction to address student deficits.

The California Standards Tests for History-Social Science, like all of the CSTs, are products of an extensive development process shared by ETS and the California Department of Education aimed at creating valid and reliable assessments. ETS has formulated testing blueprints and an “Item Utilization Plan” that guide the creation of each history test. The Department of Education and the State Board of Education approved and adopted both the blueprints and the utilization plan. The blueprints divide content standards into thematic “reporting clusters” and allocate numbers of questions across clusters. For example, the 11th grade U.S. History test includes 10 items on Foundations of American Political and Social Thought and 12 items on United States Between the World Wars.

ETS begins the development process by recruiting item writers for the California Standards Tests. All writers must have a college degree related to the history-social science “content area,” three or more years teaching experience in history-social science, and “knowledge and experience with the California content standards.”

---


28 Ibid., 82. It is important to note, however, that scale ranges vary slightly on different forms of the CSTs due to difficulty.

29 Ibid., 439.

30 In 2007, English-language arts testing accounted for 52% of middle schools’ API and 39% of high schools’ API. Math tests accounted for 34% and 28% for middle schools and high school respectively; and science tests counted for 7% and 19%. For more information on testing weights, see, California State Department of Education, Overview of California’s 2006-07 Accountability Progress Reporting System (Sacramento, CA: California State Department of Education, 2007), 3.

31 The other 11th grade clusters are Industrialization and the U.S. Role as a World Power (13 items), World War II and Foreign Affairs (12 items), and Post-World War II Domestic Issues (13 items).

32 Educational Testing Service, 98.
train question writers to follow item guidelines. These guidelines include a statement of the topic and content standard, the “expected cognitive level” of the item, the construct included in the standard, and descriptions of appropriate questions. During the construction of test items, ETS content specialists and assessment editors monitor the items that writers develop.

ETS subjects all new test items to a three part internal review. First, assessment specialists “judge the content validity of the items” to make sure that each item measures “what it was intended to measure.” To do so, they review questions against a number of criteria – for example, “the match” of items to content standards and cognitive levels, the difficulty of test questions, their “readability” and “grade level appropriateness.” After this content review, ETS puts questions through “editorial” and “sensitivity” reviews to check for “clarity” and “correctness” along with cultural “appropriateness.” ETS staff members eliminate questions considered deficient throughout the review process.

Following internal reviews, ETS presents all new test items to an Assessment Review Panel. The panel for history-social science consists of 17 members “comprised of current and former teachers, resource specialists, administrators, curricular experts, and other education professionals.” Applicants for the Assessment Review Panel are solicited from ETS and the Department of Education, reviewed by ETS Assessment Directors and approved by the State Board of Education. Review Panel members receive training from ETS on how to follow Department of Education approved guidelines for reviewing test items. The panel again checks the “match” between test questions and the measured standard and construct, the difficulty range of questions, item clarity, “correctness of answer,” “plausibility of the distracters,” and “bias and sensitivity factors.” The panel reviews all items and makes recommendations for changes. ETS then submits minutes of these meetings along with approved items to the Department of Education. Finally, in order to help the state avoid the type of controversy that surrounded the ill-fated CLAS program, the Statewide Pupil Assessment Review Panel -- a 6 member group appointed by the Governor, members of the legislature, and the State Superintendent of Public Instruction -- reviews test items to ensure that “examinees do not receive questions regarding personal beliefs or practices.”

Throughout the development and review process, ETS content and assessment specialists subject tests and questions to “classical item” and item response theory (IRT) analyses. Drawing from past test results and field-tested items, for example, they calculate point-biserial correlations – in this case, the relationship between students’ performance on a single test item with their total test score - to determine the difficulty of test questions, employ differential item functioning analyses (DIF) to examine testing differences between demographic groups, and measure test reliability. ETS claims that, overall, and in line with California’s accountability legislation, the “CSTs administered in 2007 meet the technical criteria established in professional standards for high-stakes tests, and that the items.”

Despite the considerable efforts made by ETS and the California Department of Education to establish test validity, reliability, and fairness, the history-social science CSTs appear fraught with limitations associated with multiple-
choice history tests. First, few of the questions begin to reflect the breadth of each standard's historical content. In 2005, for example, content standard 6.7.8—“Discuss the legacies of Roman art and architecture, technology and science, literature, language, and law”—was measured by the following question:

The origins of checks and balances in the United States can be traced to:

a. the French Republic  
b. the Roman Republic  
c. the Greek Aristocracy  
d. the Aztec Empire

The 10th grade test in 2006, moreover, gauged students' ability to “trace the evolution of work and labor, including the demise of the slave trade and the effects of immigration, mining, and manufacturing, division of labor, and the union movement” (content standard 10.3.4) with this test item:

In the nineteenth century, labor unions developed mostly in response to:

a. increasing unemployment  
b. government ownership of businesses  
c. wages and working conditions  
d. racial and gender discrimination

Indeed, such questions reflect slight fractions of the historical domains described by the standards. Further, where most of the content standards call for students to “analyze,” “explain,” “understand,” “discuss,” “describe,” and “examine,” many of the test questions push students to do little more than recognize and associate. Consider content standard 11.5.2:

Analyze the international and domestic events, interests and philosophies that prompted attacks on civil liberties, including the Palmer Raids, Marcus Garvey's “back-to-Africa” movement, the Ku Klux Klan, and immigration quotas and the responses of organizations such as the American Civil Liberties Union, the National Association for the Advancement of Colored People, and the Anti-Defamation League to those attacks.

40 This discussion of test items is drawn from a preliminary alignment analysis of the 195 sample questions from previous tests made available by the CDE. It is important to note that without access to any one complete test, it is impossible to know exactly how many items per test were used to measure specific standards.

41 California State Department of Education, History-Social Science Content Standards, 26.


43 California State Department of Education, History-Social Science Content Standards, 43.


45 California State Department of Education, History-Social Science Content Standards, 49.
Compare this standard with a test item that was used to measure student knowledge of it on the 11th grade CST in 2003:

Marcus Garvey’s program in the 1920s emphasized
a) vocational training
b) a back-to-Africa movement
c) integration into mainstream society
d) separate-but-equal doctrines

The apparent mismatch in cognitive skills between questions and standard is most pronounced in how the CSTs attempt to measure the Framework’s analysis skills standards. In 2004, analysis skill HR4 (Grades 10-12) – “Students construct and test hypothesis; collect, evaluate, and employ information from multiple primary and secondary sources; and apply it in oral and written presentations” – was assessed on the 10th grade test with the following item:

The streets were hot and dusty on the summer day. Stokers emerged from low underground doorways into factory yards, and sat on steps, and posts, and palings, wiping swarthy visages, and contemplating coals. The whole town seemed to be frying in oil. There was the stifling smell of hot oil everywhere.

-Charles Dickens, *Hard Times*, 1854

The historical era most likely referred to in this quotation is the:

a) Industrial Revolutions
b) Great Awakening
c) French Revolution
d) Enlightenment

These questions, like many others on the California’ s history tests, appear to illustrate what Samuel Messick referred to as “construct under-representation” and “construct irrelevant variance.” Where the former describes an assessment that “fails to include important dimensions or facets of a construct,” the latter characterizes test items that are too general and may measure something other than the target skill or understanding. Both threats to validity increase the likelihood of “false positives” where students are considered proficient on standards that they may not begin to understand.

Beyond content under-representation and the skill mismatch of test questions and standards, other troubling alignment issues characterize the relationship between the California history standards and tests. The introduction to the History-Social Science Content Standards clearly states, “The standards include many exemplary lists of historical figures that could be studied. These examples are illustrative. They do not suggest that all of the figures mentioned are required for study…” Many of the CST questions, however, feature these “illustrative” items. In 2006, for example, the 11th grade test measured content standard 11.7.3 -- “Identify the role and sacrifices of individual American soldiers, as well as the


In addition to its questionable “match” to the cognitive skill level demanded by the standard, this question appears to miss the criteria of having one clear answer. Given the chance, a student could make a case for why answers a (vocational training) and d (separate-but-equal doctrines) are also correct.


50 California State Department of Education, *History-Social Science Content Standards*, vi.
unique contributions of the special fighting forces (e.g., the Tuskegee Airmen, the 442nd regimental Combat team, the Navajo Code Talkers)\textsuperscript{51} -- with this question:

During World War II, what was the primary duty of the Navajo Code Talkers?

a. interpreting confiscated German battle plans  
b. transmitting secret messages to U.S. forces during combat  
a. translating confidential Japanese communications  
b. informing the press about the number of Allied war casualties\textsuperscript{52}

In 2004, the CST assessed 11th graders ability to “analyze the effects of and the controversies arising from New Deal economic policies and the expanded role of the federal government in society and the economy since the 1930’s (e.g., Works Progress Administration, Social Security, National Labor Relations Board, farm programs, regional development policies and energy development projects such as the Tennessee Valley Authority, California Central Valley Project, Bonneville Dam” with the item:\textsuperscript{53}

What New Deal program employed large numbers of artists and writers during the Great Depression?

a. National Recovery Administration (NRA)  
b. Agriculture Adjustment Administration (AAA)  
c. Works Progress Administration (WPA)  
d. National Youth Administration (NYA)\textsuperscript{54}

Such questions make the standards far less suggestive and much more prescriptive. By focusing on such specific content, they hold teachers and students accountable for the hundreds of figures, groups, events, and phenomena included at each grade level of the standards. Additionally, by sampling content across the standards, the tests encourage the coverage of wide expanses of history in short periods of time -- world history from 1700 to the present in 10th grade and American history from the colonial era to the present in 11th grade. The 8th grade test --which covers the 6th, 7th, and 8th grade standards – measures student knowledge of history from the Paleolithic era to the First World War.

A discussion of the consequential validity of the CSTs for History-Social Science is beyond the scope of this paper. While test scores have risen slightly over the past six years, exactly how the CSTs have influenced the teaching and learning of history in California’s public schools remains unclear. Many studies indicate “uncertain” effects of state history tests on classroom instruction.\textsuperscript{55} To date, no one has conducted such research in California. Nonetheless, the relationship between California’s history standards and tests appear at odds with the Framework’s attempt to promote “the importance of studying major historical events and periods in depth as opposed to superficial skimming of enormous amounts of material.”\textsuperscript{56} Rather than advance the ambitious teaching and learning promised by “systemic” reforms or the careful historical inquiry espoused by the history education reform movement, California’s history tests reduce the History-Social Science Framework to a long list of historical names, dates, and facts for memorization.

The California Department of Education developed the California Standards Tests for History-Social Science to help insure that significant amounts of history are taught and learned in California’s public schools. These tests have received bipartisan support and are part of an accountability system constructed over the past two decades by state Republicans

\textsuperscript{51} Ibid., 50.  
\textsuperscript{52} California State Department of Education, \textit{California Standards Test: Released Test Questions, History-Social Science, Grade 11}, 16.  
\textsuperscript{53} California State Department of Education, \textit{History-Social Science Content Standards}, 50.  
\textsuperscript{54} California State Department of Education, \textit{California Standards Test: Released Test Questions, History-Social Science, Grade 11}, 16.  
\textsuperscript{55} For example, see Grant, ed., \textit{Measuring History: Cases of High Stakes Testing Across the States}.  
\textsuperscript{56} California State Department of Education, \textit{History-Social Science Framework}, 7.
and Democrats. The Department of Education is clearly attempting to provide valid and reliable tests that measure students’ knowledge of the state’s history standards. It has invested millions of dollars into these exams and has constructed, along with ETS, elaborate procedures and multiple review processes to vet test questions. Both ETS and the Department of Education appear committed to the highest psychometric standards and make test policy and development, for the most part, transparent and open to public participation. As evidenced in the changes made to the CSTs, such as the replacement of the 9th grade test with an 8th grade test and the recent “end of course” distinction for the 10th grade test, the Department of Education has displayed flexibility in its testing policy and a willingness to make changes when necessary. All of this suggests that California’s test policies for history education are promising.

However, the CSTs for History-Social Science illustrate the problems of relying exclusively upon multiple-choice tests to measure historical knowledge. While each question does relate to the content of specific standards, they sample small fractions of the historical domains defined by the standards. Additionally, little evidence suggests that these test questions elicit and measure the higher order cognitive skills called for in the History-Social Science Framework. In this regard, it appears that ETS and the California Department of Education have set the bar low in determining the “match” between the California history standards and tests and that the CSTs suffer from questionable validity. California may have taken the lead in setting standards for history-social science education over the past two decades, but it has not equaled this commitment in devising world-class assessments for history.

Overall, the California Standards Tests for History-Social Science raise a number of questions relevant to current debates over state history assessments. For example, to what extent is it possible to determine one’s knowledge of history with multiple-choice tests? The California tests suggest that this format is limited; in fact, it is unclear exactly what types of knowledge and skills California’s multiple-choice tests measure. Constructed responses – such as short answers, historical essays, and document-based questions – are potentially more authentic and illustrative indicators of what students know about history.

Moreover, to what extent are state history assessments aligned with state standards? In addition to the weak content and construct validity of multiple-choice history assessments, the CSTs illustrate how accountability testing can alter the intent of state history standards. The History-Social Science Framework, like many state standards, is an expansive document – weighing in at just over 250 pages. The Department of Education, however, assumes the Framework will guide instruction through providing teachers with suggested historical topics and items for study. The California Standards Tests undermine this goal by turning the Framework into prescribed, rather than recommended, curriculum and therefore encouraging teachers to cover all of its content. If committed to state history assessments, we should make sure that they are aligned not only with the subject matter and skills, but also the objectives of state standards.

What are the trade-offs between using constructed response assessments and multiple-choice exams in terms of reliability and validity? Is it possible to make California’s current testing system for history-social science more valid without compromising cost, efficiency and reliability? The feasibility of developing performance assessments in a state as large, diverse, and financially constrained as California is questionable. Nonetheless, other comparable states – such as New York -- rely upon such assessments. California might benefit from examining how other states fund, develop, and conduct history-social studies testing programs that feature constructed response components.

Finally, and perhaps most importantly, what are the effects of state assessments on the teaching and learning of history? While the California Department of Education has invested millions in attempting to devise tests with adequate content and construct validity, it has yet to examine the consequential validity of these tests – a topic that remains woefully under-researched. Before deciding on whether to expand or limit state history assessments, we need to develop a clearer understanding on how such tests influence public school classrooms.

Bradley Fogo is a PhD candidate in Curriculum and Teacher Education at Stanford University. A former middle school and high school history teacher, he currently works as an instructor in Stanford’s Teacher Education Program and is a member of the Stanford History Education Group.

57 For other CSTs, ETS compares results with California’s norm referenced CAT/6 Survey tests to help determine test validity. However, there is no CAT/6 history exam, so test results of the history-social science CSTs have not been compared to any other assessment of historical knowledge. Nor has ETS subjected test items to trial think-alouds with students to help determine the cognitive processes involved in answering multiple-choice questions. For an example of such analysis, see Gabriel Reich, “Validating the Measure: State Standards, High-Stakes Assessment, and Student Historical Thinking” (paper presented at the annual meeting of the American Educational Research Association, Chicago, IL., 2007).