This is a prepublication version of the article. For the final printed version see **Breakstone, J., Smith, M., & Wineburg, S. (in press). Beyond the bubble: New history/social studies assessments for the Common Core.** ***Phi Delta Kappan*** **(Forthcoming).**

Beyond the Bubble:

New History/Social Studies Assessments for the Common Core

Joel Breakstone, Mark Smith, and Sam Wineburg

Stanford University

May 1, 2012

wineburg@stanford.edu

650-725-4411

The wait is over. The Common Core State Standards have arrived in public schools. Like a long-awaited Hollywood blockbuster, the Common Core has been the subject of intense anticipation, speculation, and scrutiny. Teachers and administrators scurried to get ready. A mini-industry of how-to-guides, curriculum maps, and professional development workshops has sprouted up. Yet, despite all this effort and the welcome focus on literacy, teachers of history/social studies still lack adequate resources to implement these standards. The biggest trouble spot is assessment.

The Common Core introduces ambitious goals for student learning. In history/social studies, students are expected to analyze primary and secondary sources, cite textual evidence to support arguments, consider the influence of an author's perspective, corroborate different sources, and develop written historical arguments -- crucial skills if students are to succeed in college and beyond. They also represent a radical turn from what was emphasized during a decade of relentless standardized testing. But if students are to master these skills, their teachers need tools to monitor growth, identify where students are having trouble, and figure out how best to help them. What tools do teachers have to do this?

Multiple-choice tests continue to dominate assessment across all subjects, but especially in history (Martin, Maldonado, Schneider, & Smith, 2011). It's easy to understand the affinity for multiple choice: they're quick and inexpensive, and the number-right score provides a seductive (if false) sense of precision. But expecting multiple-choice tests to measure sophisticated cognitive capacities is like using a pocketknife to do surgery – the tool is just too blunt. Multiple-choice questions are perhaps suited to measure aspects of factual recall, but they

are ineffective for gauging the higher-order thinking demanded by the Common Core (cf. Fredrickson, 1984; Martinez, 1999; Wineburg, 2004).

But this doesn't stop state departments of education from trying to use them, often with absurd results. Consider this standard from California's History/Social Science Framework. It asks students to "interpret past events and issues within the context in which an event unfolded rather than solely in terms of present day norms and values" (California State Department of Education, 1998, p. 41). Historians refer to this as the ability to overcome presentism (Hunt, 2002), seeing beyond our brief lifetime into the expanse of human history and how people in the past conceived of their world.

Now consider an item used to measure this understanding on California's year-end state test:

> Which was one outcome of World War II?
>
> A England and France increased their overseas possessions
>
> B The communists gained control over most of Western Europe
>
> C Japan and Germany became dominant military powers in their regions
>
> D The Soviet Union emerged as an international superpower
>
> (California State Department of Education, 2009, p. 23).

What happened to interpretation? Or placing events in context? What happened, in short, to *thinking?* If we want students to develop the skills laid out in the Common Core, it makes little sense to ask them to pick facts from a bounded list of dubious distracters.

But what are the alternatives? In history/social studies, the most highly touted one is the document-based question made famous by the College Board's Advanced Placement Program. Widely known by its acronym, the DBQ asks students to read ten to twelve documents, formulate a thesis on their basis, plan an argumentative essay, compose that essay, and then

proofread it for clarity, coherence, and correctness – all in the space of an hour. To its credit, the DBQ calls on many of the literacy skills identified by the Common Core: the ability to read multiple sources, evaluate claims, and mount arguments using evidence.

Still, given all of these moving parts, it is unclear what, exactly, the DBQ measures. Is it students' ability to engage in historical thinking and arrive at a defensible thesis? Their ability to sort through and organize disparate documents? Or their ability to express themselves in writing while wiping beads of sweat from their brows under timed conditions? Clearly, the DBQ is a worthy writing task. But is it the best tool for gauging skills like those identified by the Common Core: "attending to the ... date and origin of the information" in a source, or identifying "aspects of a text that reveal an author's point of view or purpose" (Common Core State Standards Initiative, 2010, p. 61)?

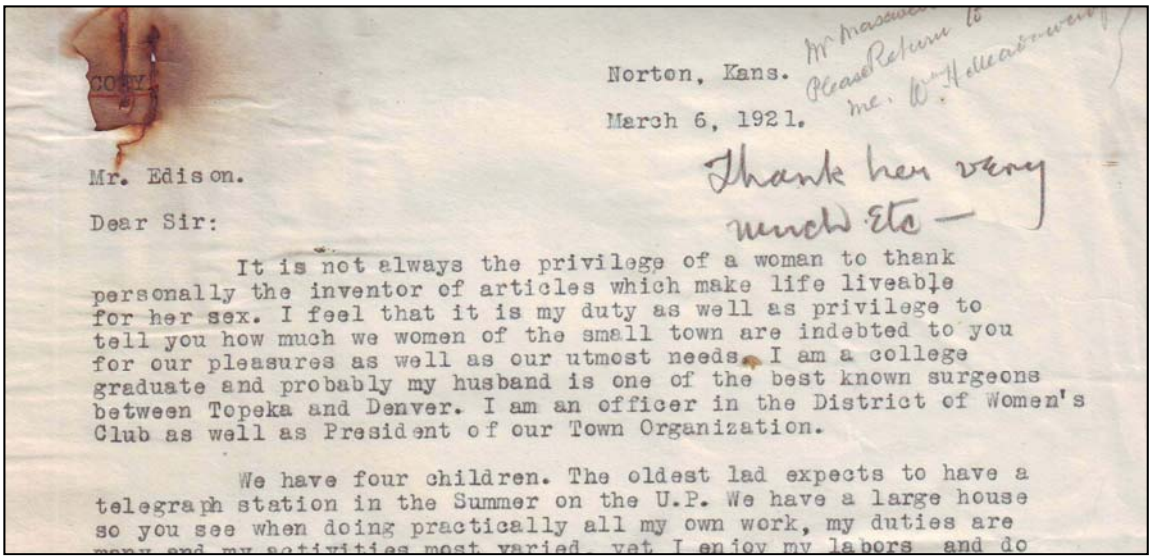
In one of the few studies that actually examined how students approached the DBQ, Katherine McCarthy Young and Gaea Leinhardt (1998) found that students often "raided" documents for appropriate quotes and facts but failed to analyze them as historical evidence. If students struggle with this college-level task, it is hard to pinpoint why -- so many things are going on simultaneously. Where are the focused assessments that can determine student needs and help them build skills to succeed on a DBQ?

History Assessments of Thinking

When we surveyed the available options, we were struck by the chasm between the rote recall demanded by multiple-choice tests and the complex orchestration of skills required by a DBQ. And, lest we forget, before students can analyze ten documents, they have to be able to analyze one. Where are the assessments for *that*?

With support from the Library of Congress's Teaching with Primary Sources Program (http://www.loc.gov/teachers/tps/), we set out to create short, focused tasks that ask students to analyze documents from the Library's vast collection of letters, books, photographs, prints, speeches, interviews, radio broadcasts, and film clips. In partnership with the San Francisco Unified School District and Lincoln (Nebraska) Public Schools, we have spent two years constructing, piloting, and revising assessments that provide teachers with new options. We call our exercises *History Assessments of Thinking*, or *HATs*. Each HAT asks students to consider historical documents and justify their answers in three to four sentences. HATs are well suited for formative assessment, one of the most effective tools for improving student achievement (Black & William, 1998). Focused assessments not only show what students are thinking, they allow teachers to locate where students are having trouble and give them ideas for what concepts to re-teach. HATs can be completed in under ten minutes, some in less than five. Even a teacher with a class of 35 students can quickly scan a set of responses to sense how well students have grasped a particular idea.

Consider this assessment, which targets a Common Core history/social studies standard: "Evaluate an author's premises, claims, and evidence by corroborating or challenging them with other information" (2010, p. 61). The task asks students to evaluate a 1921 letter written by Mrs. W.C. Lathrop, a housewife from Norton, Kansas, thanking Thomas Edison for improving her life:



COPY

Norton, Kans.
March 6, 1921.

Mr. Edison.

Dear Sir:

It is not always the privilege of a woman to thank personally the inventor of articles which make life liveable for her sex. I feel that it is my duty as well as privilege to tell you how much we women of the small town are indebted to you for our pleasures as well as our utmost needs. I am a college graduate and probably my husband is one of the best known surgeons between Topeka and Denver. I am an officer in the District of Women's Club as well as President of our Town Organization.

We have four children. The oldest lad expects to have a telegraph station in the Summer on the U.P. We have a large house so you see when doing practically all my own work, my duties are many and my activities most varied, yet I enjoy my labors and do

> It is not always the privilege of a woman to thank personally the inventor of articles which make life livable for her sex . . . I am a college graduate and probably my husband is one of the best known surgeons between Topeka and Denver . . . [Our] house is lighted by electricity. I cook on a Westinghouse electric range, wash dishes in an electric dish washer. An electric fan even helps to distribute heat all over the house. . . I wash clothes in an electric machine and iron on an electric mangle and with an electric iron . . . I rest, take an electric massage and curl my hair on an electric iron.
>
> Please accept the thanks Mr. Edison of one truly appreciative woman. I know I am only one of many under the same debt of gratitude to you.

After reading the letter, students are presented with four facts: 1) George Westinghouse invented the electric range, not Thomas Edison; 2) Before the Rural Electrification Act of 1936, less than 10% of rural America had electricity; 3) The 19th Amendment, which guaranteed women the right to vote, was passed only one year before this letter was written; 4) At the time of Mrs.

Lathrop's letter, less than 5% of American women were college graduates. While each statement is true, students must choose the two that can help them determine if Lathrop was a typical American woman of the 1920s. Unlike a multiple-choice item, students must explain their reasoning in writing -- a harder task than it might seem.

Many students have trouble figuring out which statements place Lathrop in the context of her time. Some alight on inconsequential details: "Mrs. Lathrop, who claims to have graduated from college should have known that it was not Edison that invented the electric range, but Westinghouse." Another wrote, "George Westinghouse invented the electric range, not Thomas Edison. If she was a typical 1920s woman, she would have known that. Therefore, she's atypical."

Other students are better able to set Lathrop against the backdrop of the times. As one student wrote, "Fact 4 says that less than 5% of American women were college graduates in the 1920s. Mrs. Lathrop writes in her letter that she is a college graduate, making her atypical of American women in the 1920s." Some students strengthened their answers with specific examples from the letter: "Fact 2 states that less than 10% of rural America had electricity before the Rural Electrification Act of 1936. This letter was written in 1921, which leads to the assumption that Mrs. Lathrop is atypical because she lists many examples of her use of electricity, such as an electric curling iron, electric lighting and an electric dishwasher." This student rightly questions whether Lathrop's expensive appliances were the norm in rural Kansas.

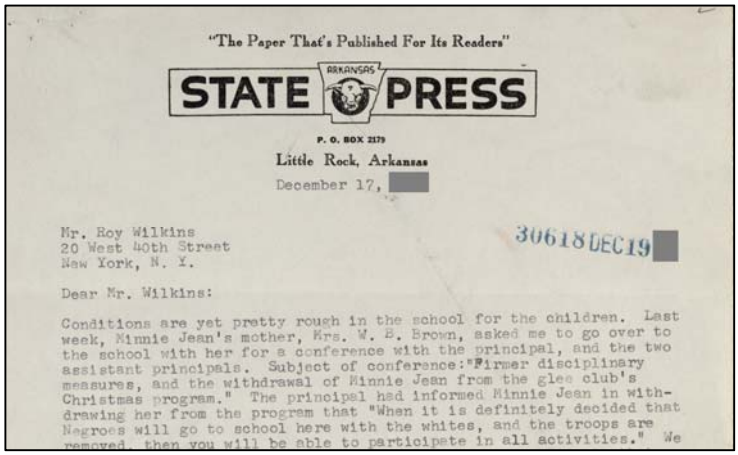
If students interpret the document through the lens of its time and place, and provide a clear rationale for their answer, teachers can move on to more complex tasks. If students struggle, their short written responses give teachers clues about where to go next.

Flexibly assessing student understanding

The letter to Edison is an example of an assessment that focuses on historical context and students' ability to make, in the language of the Common Core, "an argument focused on discipline-specific content" (Common Core State Standards Initiative, 2010, p. 64). But there are many other aspects of historical understanding. Teachers need a variety of options to monitor student progress across the full spectrum of content and skill.

Our assessments seek to address these needs. Consider a HAT that presents students with two letters drawn from the archives of the NAACP. Letter A references the President's reluctance to intervene at the state level to stop the brutal lynching of African-Americans. Letter B describes the challenges faced by African American children in a previously all-white school. The dates are removed from both letters, leaving students to answer a key question: which was written first? Instead of emphasizing the rote memorization of particular dates, this task taps into whether students can interpret documents, as well as understand key components of the Civil Rights Movement.

Even a two-line response provides a window into student thinking. Some placed letter B before letter A, arguing that the integration of previously all-white schools prompted the lynching of Black Americans by aggrieved Whites. Such a claim has a certain logical appeal. But it is wrong. These students lack an understanding of the narrative arc in the struggle for racial equality (by the time the Supreme Court ruled to desegregate schools in the 1950s, lynching had been virtually eradicated).



"The Paper That's Published For Its Readers"

STATE PRESS

P. O. BOX 2179
Little Rock, Arkansas
December 17, [redacted]

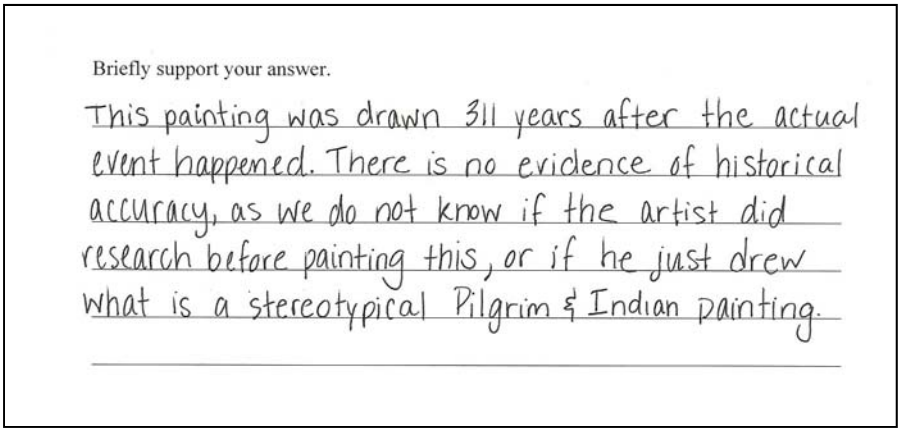
Mr. Roy Wilkins
20 West 40th Street
New York, N. Y.

30618 DEC 19[redacted]

Dear Mr. Wilkins:

Conditions are yet pretty rough in the school for the children. Last week, Minnie Jean's mother, Mrs. W. B. Brown, asked me to go over to the school with her for a conference with the principal, and the two assistant principals. Subject of conference:"Firmer disciplinary measures, and the withdrawal of Minnie Jean from the glee club's Christmas program." The principal had informed Minnie Jean in withdrawing her from the program that "When it is definitely decided that Negroes will go to school here with the whites, and the troops are removed, then you will be able to participate in all activities." We

A different type of HAT addresses a Common Core expectation that students will consider a document's date and origin when making judgments about its trustworthiness (Common Core State Standards Initiative, 2010, p. 61). Students are presented with an image of the first Thanksgiving, painted in 1932, and must explain whether it would be useful to historians trying to reconstruct relations between Pilgrims and Indians in 1621. A 311-year gap separates the painting from the event. Yet many students skip over this information entirely. Rather than considering the three intervening centuries, ample time for distortions, myths, and legends to seep into collective memory, many students focus exclusively on the painting's rich details, never considering its attribution. One wrote, "You can see how they are interacting with each other. Without any picture, you couldn't really see how Wampanoag Indians and the Pilgrims acted." Other students, however, demonstrated a firm understanding of the importance of a document's date: "This painting was drawn 311 years after the actual event happened. There is no evidence of historical accuracy, as we do not know if the artist did research before painting this, or if he just drew what is a stereotypical Pilgrim and Indian painting." In both cases, the students' written responses provide teachers with information that informs future instruction.


Briefly support your answer.

This painting was drawn 311 years after the actual event happened. There is no evidence of historical accuracy, as we do not know if the artist did research before painting this, or if he just drew what is a stereotypical Pilgrim & Indian painting.

We know that effective formative assessment requires continually monitoring student progress. If students do not master a particular concept, teachers can revisit it to assess students

again. To do this well, students will need multiple versions of the same type of assessment. To that end, and to give teachers maximum flexibility, we have created parallel versions of each HAT which contain documents from different eras.

Conclusion

We have long understood that the form and content of tests profoundly influence the type of classroom instruction that students receive (Frederiksen, 1984; Madaus, Russell, & Higgins, 2009). If we want students to achieve the benchmarks set out in the Common Core Standards, then we need assessments that are aligned to these skills. The educational community has shown that it can produce high-quality standards documents that lay out inspiring and worthy educational goals. But without concrete tools that assess student progress toward those goals, this new round of standards, like previous rounds, may founder on the shoals of rhetoric and verbiage. HATs will not solve this problem. But they may help ignite our creativity so that we can develop effective, efficient, and worthy tools for assessing student understanding.

The Stanford History Education Group's HATs are freely available on a new website (beyondthebubble.stanford.edu). In addition to offering these assessments, the site features annotated samples of student work and easy-to-use scoring rubrics. There are also short videos with tips for teaching with historical sources and implementing HATs.

This work is generously supported by the Bill and Melinda Gates Foundation and the Library of Congress's Teaching with Primary Sources Program. However, no endorsement of the views expressed here should be inferred from this support.

References

Black, P., & Wiliam, D. (1998). Inside the black box: Raising standards through classroom assessment. *Phi Delta Kappan, 80*(2), *139-144.*

California State Department of Education. (1998). *History-social science content standards for California public schools*. Sacramento: California State Department of Education.

California Department of Education. (2009). *California Standards Test: Released test questions World History.* Sacramento: California State Department of Education.

Frederiksen, N. (1984). The real test bias: Influences of testing on teaching and learning. *American Psychologist*, *39*(3), 193-202.

Hunt, L. (2002). Against presentism. *Perspectives of the American Historical Association.* Retrieved from http://www.historians.org/perspectives/issues/2002/0205/0205pre1.cfm

Madaus, G., Russell, M., & Higgins, J. (2009).*The paradoxes of high stakes testing*. Charlotte, NC: Information Age Publishing.

Martin, D., Maldonado, S. I., Schneider, J., & Smith, M. (2011). *A report on the state of history education: State policies and national programs*. National History Education Clearinghouse. Retrieved from http://teachinghistory.org/system/files/teachinghistory_special_report_2011.pdf

Martinez, M. E. (1991). A comparison of multiple-choice and constructed figural response items. *Journal of Educational Measurement, 28*(2), 131-145.doi:10.1111/j.1745-3984.1991.tb00349.x

Wineburg, S. (2004). Crazy for history. *Journal of American History*, *90*(4), 1401-1414.

Young, K. M., & Leinhardt, G. (1998). Writing from primary documents: A way of knowing in history. *Written Communication*, *15*(1) 25-68.