“Testing must be a tool in our toolbox, but we need more rulers and fewer hammers,” wrote Richard Woods, Georgia’s School Superintendent, in a recent letter to Secretary of Education Arne Duncan. Instead of using high-stakes tests to penalize schools, Woods called for a “formative assessment model that identifies the strengths and weaknesses of students.”

What is formative assessment? How does it differ from its more familiar cousin, summative assessment? A leading expert on assessment, Dylan Wiliam, explained the difference:

Traditional approaches to instruction and assessment involve teaching some given material, and then, at the end of the teaching, working out who has and hasn’t learned it—akin to a quality control approach in manufacturing. In contrast, assessment for learning involves adjusting teaching as needed while learning is taking place—a quality assurance approach (p. 19).

The key to understanding formative assessment is the subtle but all-important difference between the “for” and the “of” that joins “assessment” to “learning.” Rather than using a test to put a grade on a students’ permanent record, formative assessment is used to inform teaching—to give teachers usable information about what students know and what they are still struggling to understand. In a comprehensive review of 250 studies, Dylan Wiliam and colleague Paul Black found formative assessment to be one of the most
powerful tools a teacher has to boost student achievement. The benefits of formative assessment were strongest for struggling students.ii

Despite these proven benefits, formative assessment remains rare in American classrooms. The very concept goes against what David Tyack and William Tobin called the “grammar of schooling,” where assessment is synonymous with standardized testing.iii But as Rick Stiggins noted, standardized tests “cannot inform the moment-to-moment, day-to-day, and week-to-week instructional decisions faced by students and teachers seeking to manage the learning process as it unfolds.”iv Assessment for learning—if it is to mean anything—must provide teachers with information that helps them actually change how they teach.

Compared to the vast sums poured into standardized testing, formative assessment is a poor stepchild.v A 2006 report from UCLA’s National Center for Research on Evaluation, Standards, and Student Testing concluded that teachers “lack the types of systematic and sensitive assessments [they] need to . . . make visible students’ thinking.”vi To fill this gap, some districts turn to multiple-choice tests. But blackened circles defy the purpose of formative assessment: to provide insight into students’ thinking. A blackened mark can mean any host of things—that the student got the wrong answer for the right reasons, the right answer for the wrong reasons, or simply marked C because there had already been too many As and Bs.vii Furthermore, as the University of Colorado’s Lorrie Shepard cautioned, using multiple-choice items for formative assessment leads to the “1,000 mini lessons problem.” To address all of the items students got wrong, teachers would have to teach an endless series of short lessons, resulting in an “incoherent and decontextualized” curriculum.viii
At the other end of the spectrum from multiple-choice tests is the document-based question (DBQ). In its best-known version, students read eight to 12 sources, formulate a thesis, and incorporate the documents into an argumentative essay—all in less than an hour. The DBQ is a rich task but one ill-suited for frequent cycles of assessment. The time it takes students to complete DBQs pales in comparison to how long it would take teachers to grade multiple DBQs each week from the 150 - 200 students they see every day.

**A Different Approach**

In searching for alternatives to multiple choice tests and DBQs, we were inspired by the common practice of “do-nows” (also known as “bell work”) in which teachers give students a brief task at the beginning of class to prepare them for the day’s lesson. Could these minutes at the start of class be used for formative assessment?

With support from the Library of Congress, we set out to create tasks that would give social studies teachers new options. Our exercises are built around primary sources from the Library’s extensive collection. Each exercise assesses core skills of historical thinking and provides teachers with quick feedback about student understanding. All of our tasks, which we call History Assessments of Thinking (HATs), can be completed in less than ten minutes, some in less than five. They require students to evaluate a source and to write a few sentences to explain their reasoning about it. Even short answers, we have learned, tell us a great deal about student thinking.

Consider a task in which students are presented with two letters written to the executive secretary of the National Association for the Advancement of Colored People (NAACP). The first describes the reluctance of the president to intervene in the “lynching
situation” because it was a “question of education in the states.” In contrast, the second letter describes violence against African-American children in a previously segregated school, and implores the president to intervene. Both letters appear without dates and students are asked a simple question: Which letter came first? Rather than recall the date of Brown v. Board or identify when lynching reached its apex, students must use historical knowledge to place these documents into the narrative arc of the Civil Rights movement. They must then justify their answer in a short written response.

Other assessments focus on particular aspects of historical thinking, such as who wrote a document, when, and why (“sourcing”). One exercise targets students’ ability to recognize the gap between when a source was created and the event it describes. Students are presented with an 1859 engraving showing the 1620 signing of the Mayflower Compact and asked whether this source would be useful to historians seeking to understand this event. Many students fixate on the content of the image, ignoring the image’s date entirely. Stronger responses address the limitations of a source created more than two centuries after the event it depicts.

Another task asks students to consider how a document’s context shapes its content (“contextualization”). Students are presented with Dorothea Lange’s iconic “Migrant Mother” photograph of Dust Bowl migrants in California (see Figure 1). They are also provided background information about Lange’s employment with FDR’s Resettlement Administration, and her charge to take photos that would build support for his social programs. Students are asked why such information—the fact that Lange had a particular motivation for taking this picture—might lead to questions about its reliability.
We reviewed thousands of responses from middle and high school students and found that students’ answers—even if only in a brief sentence—open a window to how they think. Many students struggled to understand the photo in context (see Response 1 in Figure 2). Those that used the background information sometimes concluded that the picture was suspect and not useful as evidence about the Dustbowl (see Response 2 in Figure 2). But can Lange’s photo be dismissed so easily? To probe students’ thinking further, our exercise asked them what else they would want to know about Lange (or how she took the picture) in order to better evaluate the photo’s reliability.
We drafted dozens of other HATs, piloted them with students across the country, and revised them until we were confident that they gauged specific aspects of historical thinking. The tasks that ultimately survived this process are freely available at beyondthebubble.stanford.edu, a website that also features rubrics, sample student responses annotated to explain common misconceptions, streaming videos, and links to the original documents in the Library of Congress’s digital archive. We also created multiple versions of each HAT. A parallel version of the Lange task provides students with an image of a child mine worker and background information explaining that Lewis Hine, the photographer, had been hired by the National Child Labor Committee. These alternate versions, using different content, allow teachers to see if their students have grasped the underlying historical concept.
How might busy teachers use our new assessments? To find out, we observed a group of high school teachers from San Francisco public schools who agreed to try them out in their classrooms. These teachers had worked with us during the development of our Reading Like a Historian curriculum, and were committed to helping students master the elements of historical thinking.

We quickly discovered that notions of assessment were deeply ingrained for teachers and students. Teachers felt compelled to assign scores to students’ responses and students expected to receive grades. As one of the teachers, Chris, explained, “Students feel like they need a grade, and then I feel like, Oh, I need to put this in a gradebook . . . that is an overwhelming feeling sometimes, but what [students] really take out of a classroom is what counts, right?” Researchers have called this focus on grades the “summative assessment teaching script.” For the teachers we worked with, long-held beliefs about assessment sometimes led to a “rush to rubrics” in order to assign grades. At the beginning of our work together, teachers sometimes spent more time categorizing student responses than reflecting on what those answers might tell them about student thinking.

At the same time, we were heartened when teachers used our tasks in ways we didn’t imagine. For example, Seth, a five-year veteran of the social studies classroom, gave students the Migrant Mother exercise and graded them using our three-level rubric. He returned students’ papers the next day. But rather than moving on to a new topic, he used the next class to display exemplary responses and to talk about how he graded them. As Seth went through the answers, he asked students to revise their answers using a different colored pen. Seth demonstrated for students the types of responses he expected
and provided them with an opportunity to improve their answers. In the process, he not only received feedback about his students’ growth as historical thinkers but immediately gave students a chance to address shortcomings in their understanding.

Teachers also used HATs to get students to grapple with the nuances of historical thinking. Drawing on the information about Dorothea Lange’s employment, many students rejected her photo as untrustworthy. Students had trouble understanding that they could both question Lange’s motivations and identify valuable aspects of the photo as evidence. This type of dichotomous thinking troubled the teachers—interpretations of sources are rarely so cut and dry. Teachers wanted students to contend with the shades of gray that invariably accompany the evaluation of documentary evidence. To complicate students’ thinking, teachers showed students the unpublished photos that Lange took of the same family (see Figure 3). When students objected to the photos as staged, Seth asked: “Who has ever posed in a photo? Everyone should put a hand up because you take school pictures every year. Are those photos not reliable?” Similarly, Seth’s colleague, Amy, an 11-year-veteran, asked her now-suspicious students to give Lange’s photo a second look, “You don’t think the lines in her face tell a story, too?” Here, the process of formative assessment, which initially addressed the context surrounding a photo, sparked a conversation about the inferences that can be drawn from fragmentary evidence. Instead of memorizing someone else’s conclusions, students became active participants in a debate about the very question of what it means to do history.
During the time we worked with teachers, we noticed shifts in their practice. Rather than focusing primarily on assigning grades, teachers began to consider how to use our assessments to foster student thinking. During one afterschool meeting, Amy suggested the possibility of assigning a task and not grading it all: “I mean, technically, couldn’t you give this—collect everybody’s response and then put it up there and be like—okay, so what do people do? Go over it right there?” Eventually, all of the teachers incorporated some version of this idea into their practice.
Despite the brevity of our tasks, it was still time consuming for teachers to review student responses. Seth noted that grading students’ short answers to the Lange task had still taken several hours: “It was a great assessment, but I have a hundred students . . . When you’re grading a hundred, there’s always the question of, How much feedback can you give and be timely about it?” Seth rightly noted that delays in responding to students compromised the purpose of formative assessment: “If they don't get it back quickly, it's less formative and more summative.” This tension persisted throughout the time we worked with these teachers.

**Formative assessment in the social studies**

We were impressed by the creative ways in which our teacher colleagues adapted our tasks to foster students’ thinking. We also realize that more support is needed to make formative assessment a regular part of the social studies classroom. From the Library of Congress’s unparalleled archives, we have created a collection of assessments that can aid the move toward assessment *for* learning. But more is needed. Stakeholders from across the spectrum must come together to effect meaningful change. Policymakers like Secretary Duncan must act on Superintendent Woods’s call for a shift toward formative assessment. Duncan has acknowledged as much. In 2013, he told a group of educational researchers,

> Not enough is being done to use high-quality formative assessments to inform instruction in the classroom on a daily basis. Too often, teachers have been on their own to pull these tools together—and we've seen in the data that the quality of formative tools has been all over the place.
This work is not only the responsibility of policymakers. Test developers and researchers—ourselves included—need to do more to ensure that the assessment toolbox is stocked with classroom-ready tasks, particularly ones linked to specific lessons. But even such curricular resources won’t be enough. We need to construct digital platforms to streamline the process of formative assessment. After reading thousands of student answers to the Lange HAT, we are familiar with the most common student responses. Our San Francisco colleagues could have saved precious hours had that knowledge been put into a digital format that facilitated easy feedback to students.

These are the challenges that await us. Instead of conceiving of assessment as a terminus, something to do at the end of a unit or course, it must become a routine part of teaching. Admittedly, this shift won’t be easy. But if we can disrupt long-held beliefs about assessment as an end product, we can guarantee better outcomes for us and for our students.

---


James C. Curtis, “Dorothea Lange, Migrant Mother, and the Culture of the Great Depression,” *Winterthur Portfolio* 21, no.1 (1986): 1-20; Erroll Morris, “The Case of the Inappropriate Alarm Clock,” *The New York Times* (October 21, 2009), opinionator.blogs.nytimes.com/2009/10/21/the-case-of-the-inappropriate-alarm-clock-part-4/; James Curtis, professor of history at the University of Delaware, used the additional photographs that Lange took to build an argument that Lange had carefully choreographed the photographs to ensure an image that would best serve the purposes of her employer, the Resettlement Administration. In 2009, documentary filmmaker Erroll Morris challenged this narrative in a piece written for the *New York Times*. Morris wrote, “Is the argument that [Lange] should not have intended to produce any effect? None whatsoever? . . . Isn’t the issue a deeper, far more problematic issue, namely, did the photographer intend to deceive the viewer? Was he or she using photography as sleight of hand? Is the problem, intentions? But if so, how could one do away with them altogether? We read correctly or incorrectly a photographer’s intentions into every photograph we see” (para. 21). Although none of the teachers mentioned Curtis or Morris by name, their lessons allowed students to engage in this same historical debate.