

# Measuring College Learning in History

**Lendol Calder**  
Augustana College

**Tracy Steffes**  
Brown University

*This contribution advances a case for why historians must come together not only to articulate the value of historical study but also to demonstrate its value with evidence. The authors argue that today's students should develop a deep understanding of history as an interpretive account, the relationship of past and present, historical evidence, complex causality, and historical significance. In addition to mastering these essential concepts, today's history undergraduates should learn how to evaluate historical accounts, interpret primary sources, apply chronological reasoning, contextualize, and construct acceptable historical accounts. Following their in-depth discussion of learning outcomes, the authors review existing history assessments in K–12 and higher education. These include well-known tests like the Advanced Placement history tests and newer tools such as the Stanford History Education Group's Beyond the Bubble assessments. The authors conclude with a vision for the future of assessment in the discipline of history.*

## Introduction

In 2004, Richard Rothstein published an article in the *Journal of American History* titled “We Are Not Ready to Assess History Performance.” A former national education columnist for the *New York Times*, now a respected education analyst for the Economic Policy Institute, Rothstein argued confidently that large-scale, standardized history assessment is impossible. It cannot be done, Rothstein concluded, for the simple reason that no public consensus exists about what history students should learn. Some want history instruction to foster American national identity with stories of heroes and triumphs. Others want to prepare the young to fight for social justice by learning about the power structures that benefit some while oppressing others. Still others want to foster civic and cultural literacy. To be meaningful, large-scale assessments must be aligned with accepted objectives for learning. If Americans do not agree on the outcomes for history education, Rothstein’s conclusion follows: “This renders standardized assessment impossible” (2004, 1390).

Since 2004, periodic public controversies over history instruction show that Americans continue to disagree about its aims, especially the key goals, content, and narratives to teach in K–12 schools. For example, in 2014, when the College Board released a revised curriculum framework for Advanced Placement (AP) U.S. history, the new framework garnered praise from the American Historical Association (AHA) and the Organization of American Historians (OAH). But conservatives denounced the new course for being insufficiently celebratory about the nation’s past and radically revisionist. The new AP framework was censured by school boards, threatened with defunding by a handful of state legislatures, and described by Republican presidential candidate Ben Carson as so “‘anti-American’ that most people completing the course will be ready to sign up for ISIS” (Lerner 2015).

Implicit in such critiques of revisionist history is a belief shared by many Americans—perhaps most—that history is what really happened, that is, a single, so-called right story of settled truths. In Florida this view is written into state law. A 2006 statute to raise historical literacy requires the state’s public school history teachers to limit themselves to the “teaching of facts,” stipulating that “American history shall be viewed as factual, not constructed . . . and shall be defined as the creation of a new nation based largely on the universal principles stated in the Declaration of Independence” (Florida K-20 Education Code 2015, 1003.42). On this view of the matter, it follows that achievement in history is to be measured by students’ ability to remember and reproduce an authorized, unchanging canon of important facts and stories.

Yet for most college history faculty, history is all about interpretation. Historians are likely to agree with R. G. Collingwood, author of the canonical *Idea of History*, that “nothing capable of being learnt by heart, nothing capable of being memorized, is history” (1939, 75). Historians think of history not as settled truths about the past but as a sense-making activity, always and inescapably interpretative. History is a constructed explanation made from fragmentary evidence that is always incomplete and subject to revision. Constructing these narratives, historians draw from a discipline situated at the intersection of humanities and social science, with a distinctive set of ideas and practices that guides how we interpret and use evidence, weigh and evaluate plausible explanations, and think about the past in relation to the present. Thus, to historians, popular notions about history teaching largely misunderstand what is most valuable and meaningful about history as a discipline and misrepresent the goals of history teaching and learning, especially at the college level.

Returning to Rothstein’s point, it is true that historians and the public do not see eye to eye about outcomes for history education. But the *we* in Rothstein’s conclusion that “we are not ready to assess history performance” refers to the adult stakeholders of

K–12 education: parents, teachers, lawmakers, and policymakers. If we turn from the K–12 scene to higher education, different conclusions about assessment in history become possible and, we argue, necessary.

At the college level, historians disagree among themselves about certain aspects of history teaching and learning, such as how to balance breadth and depth, for example, and whether to require the study of any particular histories. But for the most part college history instructors share commitments to a disciplinary set of norms and assumptions that structure our goals and approach to teaching. In the past two decades, historians and others who have studied these disciplinary norms have documented a considerable amount of agreement among historians when compared with discussions about K–12 history education. Nearly all the scholarship and faculty-led efforts to articulate history learning outcomes assume that college students should learn about history *as a discipline*, with emphasis on the habits of mind of historical thinking. Historical thinking honed through disciplined study of the past offers many benefits for students. It fosters critical thinking and analysis skills that are useful for work, citizenship, and individual efficacy. History offers a critical perspective on the present and satisfies a natural longing most people have to situate themselves in a larger context and stream of time. A historical consciousness fosters perspective taking and empathy, and, because it requires students to wrestle with the limits of knowledge, historical thinking is a training ground for solving problems when definitive answers are elusive.

Still, despite the agreement among college history faculty about the goals and purposes of history instruction framed around disciplinary norms, few have stepped forward to challenge Rothstein's conclusion that standardized assessment is impossible. This is not surprising. As experts in the discipline, many of us have developed assessments tailored to individual courses that evaluate the learning outcomes we care most about. We have faith that the accumulation

of experiences across history classes produces the student learning we desire (as it did for most of us). We value our autonomy and may view standardized assessment tools as a threat to it. Furthermore, news headlines given to K–12 testing regimes and burgeoning conversations about standardized assessment in higher education do not inspire enthusiasm for new assessment tools. Many faculty worry, not without reason, that the push for assessment will saddle faculty with irksome measures poorly designed to capture meaningful student learning or, at worst, measures that degrade history teaching and learning.

We share these concerns. No one wants ill-advised assessment regimes imported into higher education. No one wants to see a single-minded, narrow emphasis on quantifying value. No one desires deeply flawed metrics being used to compare institutions and individuals. Nevertheless, we believe it would be a serious mistake to let dissatisfaction with existing assessment lead historians to eschew all forms of evaluation or to refuse to engage in conversations about how to measure learning. In light of trends already reshaping higher education, we argue that cynicism about assessment is dangerous for the history profession. If historians do not come to the table for conversations about assessment, decisions will be made without us. The risk we take in opposing all forms of assessment is letting policymakers or external authorities impose on us tests that are far less valid and useful than ones we might have designed ourselves.

As historians, we live in a moment of declining history enrollments, popular attacks on the humanities, and growing demands that a college education have practical utility and demonstrable economic benefits. Decades of rapidly escalating costs for higher education have brought growing scrutiny from external stakeholders. With a majority of Americans believing that higher education in the United States is not providing students good value for their tuition dollars and an even larger majority—75 percent—saying college is too expensive to afford, policymakers and the public

are pressing colleges and universities, and fields of study within them, to demonstrate their value (Taylor et al. 2011). Responding to this pressure, Mitch Daniels of Purdue University asked faculty to develop metrics to better measure learning outcomes (Flaherty 2015). The situation at Indiana is a sanguine version of a future awaiting the rest of us; other presidents may not seek faculty input. In an environment like the present, a rigid opposition to assessment is not an option.

If historians should not reject standardized assessment out of hand, neither should we view history assessment as simply a defensive move. In its worst forms, cynicism about assessment is indistinguishable from anti-intellectualism about history teaching and learning. As one tool in a larger assessment toolbox, a rigorously developed standardized instrument designed by historians could help us gather important evidence about student learning according to the criteria that we as historians deem most important for the discipline. It could help us to make more informed decisions about our teaching and curriculum and to explore with more evidence and precision what our students are learning. It could also help us make a stronger case to external audiences including students, university administrators and accreditors, employers, and policymakers about the value of history by helping to *demonstrate* the transferable knowledge, skills, and habits of mind history majors carry with them into the world and the ways historical study empowers them as citizens, workers, and individuals. A standardized assessment for college-level history could perhaps alter public perceptions about what history is for. The effort is worth making because now, as never before, the gulf in understanding about history between disciplinary experts and much of the public matters.

Historians today face a momentous opportunity to articulate clearly and persuasively the value of historical study for college students, most of whom will not pursue graduate study or careers in the historical professions. The AHA, through its Tuning Project, has begun this work already, supporting history departments across the

nation to articulate the learning outcomes of a history major and communicate its value to students, employers, and policymakers. Our aim in this white paper is to advance this conversation already occurring within the discipline and to take it a step further. We want to explore how we can move from asserting the value of historical study to demonstrating it with evidence. Our conclusions reflect the perspective of disciplinary experts—historians and college history faculty—about what it means to be good at history and how we can measure student performance on tasks requiring a historical eye.

We begin by surveying the development over time of efforts to define essential disciplinary concepts, competencies, and habits of mind. We use these past and current conversations, along with insights and feedback from the diverse members of the Measuring College Learning (MCL) in History faculty panel, to construct a list of *essential concepts and competencies* that can serve as the basis of assessment in history.<sup>1</sup> This list is meant not to be comprehensive or exhaustive but rather to identify a focused set of essential concepts and competencies that have broad agreement within the discipline as fundamental, important, and valuable goals for history majors. We do not say these are the only things worth assessing in history or that this should be the only assessment—there should always be multiple measures—but this provides one way to assess important, fundamental, and valuable historical learning outcomes. After defining and justifying these concepts and competencies, we discuss current methods of history assessment and

---

<sup>1</sup> In addition to the white paper authors, the history faculty panel included Julia Brookins (American Historical Association); Elaine Carey (St. John's University); James Grossman (American Historical Association); Anne Hyde (Colorado College); Norm Jones (University of Utah); Kenneth Pomeranz (University of Chicago); Nancy Quam-Wickham (California State University, Long Beach); Maris Vinovskis (University of Michigan); and Emily Swafford (American Historical Association).

propose some ideas for what an assessment framed around these essential concepts and competencies might look like. We end by considering potential uses for a standardized history assessment tool and suggesting ways that the profession can engage in conversations that will extend our collective vision of the possible in the realm of valid, useful history assessment.

By itself, the idea of assessment is not difficult for historians. In our courses, we tie assessment organically to teaching as an act of historical inquiry into learning: all assessment is, in essence, a historical argument about something (learning) that happened. This white paper explores how we might do this at the level of the history major. All historians believe passionately in the value of our discipline. Now more than ever we need to communicate this value with tools that capture useful information about student learning.

## A History of History Learning Outcomes

College history teachers have only recently adopted the terminology of *learning outcomes*. However, efforts to specify what history students should know, do, and value are not new at all but span back over a century with contributions from a number of disciplinary perspectives. Philosophers, historians, history educators, and recently cognitive scientists have applied themselves to the problem of defining historical understanding and best practices for history education. Today these braided conversations inform newly invigorated efforts of history faculty to theorize history learning outcomes and how we know whether students meet them. Using very broad strokes, we document here how thinking about the nature and purposes of history has been shaped over time by larger developments in the history profession and in K–16 education.

One important body of work affecting ideas about history teaching is that produced by philosophical reflection on the nature



and meaning of historical consciousness. Exemplars include explorations written more than a half century ago by R. G. Collingwood, Marc Bloch, and Edward H. Carr, books that have the status of classic works today (Bloch 1953; Carr 1961; Collingwood 1946). Their reflections on the theory and practice of history have recently been supplemented by efforts to define what is distinctive and valuable about the discipline of history and defend the historian's craft from postmodern theories on one hand and social science critiques on the other (Evans 2000; Gaddis 2004). Of course, interpretative debates within the field have been many, varied, and at times vociferous. The magnitude of the quarrels is described by Keith Jenkins:

Would you like to follow Hegel or Marx or Dilthey or Weber or Popper or Hempel or Aron or Collingwood or Dray or Oakeshott or Danto or Gallie or Walsh or Atkinson or Leff or Hexter? Would you care to go along with modern empiricists, feminists, the Annales School, neo-Marxists, new-stylists, econometricians, structuralists or post-structuralists, or . . . Marwick . . . , to name but twenty-five possibilities? And this is a short list! (Jenkins 1991, 18)

These interpretative debates involve substantial disagreements about what questions to ask, which voices and phenomena to prioritize, and which kinds of explanations are most persuasive. If philosophical reflection upon historical consciousness and methodology were all one had to go on for delineating history learning outcomes, we might conclude that consensus on disciplinary goals and methods is unattainable.

However, a second body of literature offers grounds for hope. History primers—that is, books, articles, and handbooks written for or about the training of history students—suggest that, notwithstanding their different schools and ideologies, historians

share conceptual reference points that, once identified, offer clear markers to delineate competence in history. By identifying and demystifying the key intellectual “moves” and understandings common to expert historians, primers aim to foster student learning in history courses. What are the concepts and competencies said to be the common property of most historians and important enough to teach to undergraduates?

The short answer is that our understanding of these key moves has changed over time. For example, historians have long debated what outcomes realistically can be expected in the introductory college course. The nature and extent of this debate, and what it reveals about college faculty’s expectations for history learning, are described by Joel Sipress and David Voelker in a 2011 essay examining how introductory courses became surveys of broad historical knowledge. According to Sipress and Voelker, the educational goals and methods defining the traditional introductory history course were settled over a century ago when coverage of historical information was made the goal of the introductory course and cultural literacy the chief learning outcome. Coverage and cultural literacy did not go unchallenged. Sipress and Voelker document moments in every generation of historians when reformers criticized the coverage model. For example, at the December 1897 meeting of the AHA, when historians debated whether or not primary source materials had a place in introductory courses, the University of Nebraska’s Fred M. Fling argued for making source work “the staple of historical instruction” (Sipress and Voelker 2011, 1054). The following year, a Committee of Seven appointed by the AHA to study the issue agreed with Fling, contending “that the accumulation of facts is not the sole, or perhaps not the leading, purpose of studying history” (Sipress and Voelker 2011, 1054). Yet the recitation of authoritative knowledge continued to be the goal of a coverage-oriented history teaching that retained its dominance through the 20th century and persists today. Sipress and Voelker argue this was the case because the psychology of the coverage

model was reinforced by historically contingent trends shaping higher education, including assumptions about (a) learning (e.g., early 20th-century behaviorism and its attic theory of cognition in which the mind must be stocked with facts before one can learn to think about them); (b) the purpose of history (e.g., the view, popular after WWI, that the primary burden of history instruction was the formation of citizens who would be safe for democracy, or know the right things); and (c) general education (e.g., the view after WWII that democracy was threatened by deficiencies in what students knew about Western and American civilization) (Sipress and Voelker 2011).

In recent decades, new challenges to the coverage model are undermining its hegemony and suggesting different outcomes for history education. An explosion of historical knowledge and subfields makes coverage more than ever an impossible objective. Even more damaging are the rejection of behaviorism in cognitive science and reforms in K–12 education that turned the attention of educators from teaching to learning. After *A Nation at Risk* (1983) warned the country was facing an educational crisis wrought by a “rising tide of mediocrity,” education reformers embarked on a wide range of efforts to raise educational standards, frame ambitious curriculum content and student performance standards, and assess students’ and schools’ progress toward meeting those goals (National Commission on Excellence in Education 1983). Many supporters viewed standards-based reform as an equity effort, since it focused attention on whether students were learning and pressured schools to ensure that they did. Since the No Child Left Behind Act in 2001, this pressure on schools has become the dominant attribute of federal policy, a pressure that manifests itself through a high-stakes testing regime that issues escalating school sanctions based on test performance in the name of accountability. This testing has spurred strong and growing resistance because of its equity implications, negative impacts on teaching and learning, and simplistic equation of test scores on low-level standardized tests with school quality.

The backlash against this testing threatens to undermine assessment more generally, despite the conviction of many education reformers that appropriate assessments, tied meaningfully to curriculum, can help to enhance quality and equality in education by focusing attention on student learning.

In history, the Bradley Commission on History in Schools, which included over a dozen eminent historians along with teachers and writers, stepped into this national reform conversation about curriculum goals and standards. It articulated a rationale for studying history in schools, core themes and narratives to emphasize regardless of content, and elements of historical thinking. The Bradley Commission's *Building a History Curriculum* (1989) influenced the National Standards for History undertaken a few years later as part of an effort to develop voluntary national standards in core school subjects. After becoming embroiled in mid-1990s culture wars over what students should know and whether the standards were sufficiently celebratory of the nation's past and European tradition, the effort to create national standards was abandoned. In the rancorous debate about what content students should know, some of the innovations of the effort were buried, especially its emphasis on analytic thinking and skill-building activities to make history come alive in the classroom. However, these innovative efforts continued and gained ground in state-level curriculum standards reform, educational scholarship and teacher education, and some large-scale assessments like NAEP and the Advanced Placement history examinations.

For the most part, college history faculty members stand aloof from these changes and discussions. Yet as Sipress and Voelker relate, seeds of reform planted in the 1990s now promise to deliver the long-sought goal of pedagogical reformers: the end of the coverage model. One source of change was innovative theories and practices percolating up from K–12 history education, often endorsed by historians active in the preparation of future school history teachers. Another was the Carnegie Foundation's sponsorship

in 1999–2000 of a cadre of historians to jump-start the scholarship of teaching and learning (SoTL) in history. Early on, history SoTL specialists realized that one of the most important tasks in front of them was replacing the coverage model's understanding of what it means to be proficient at history—which unintentionally reinforced public misperceptions of history as important things that happened—with new understandings of expertise based on how historians think and tuned to how people learn. To the early SoTL scholars, this was interesting intellectual work on its own terms, benefiting students and historians alike (Calder, Cutler, and Kelly 2002; Pace 2004). Only later, as calls for accountability in higher education grew louder, did it occur to anyone that there might be other uses for specifying expertise in history and designing valid assessments, such as defending history's value in a liberal arts curriculum.

In the last two decades, then, numerous workbooks, guides, and articles by SoTL scholars have defined and argued for the core concepts and skills that should be at the center of history teaching. On this question agreement in the literature of history primers is not perfect, but it is substantial. Conal Furay and Michael Salevouris's *The Methods and Skills of History: A Practical Guide* (1988) proposes five core elements of historical-mindedness: sensitivity to how other times and places differ from our own; awareness of basic continuities in human affairs over time; ability to note and explain significant changes; sensitivity to multiple causation; and awareness that all written history is reconstruction that inadequately reflects the past as it really happened. Thomas Holt (1995) identifies two characteristics of historical-mindedness every college student should master: analytic questioning and the ability to synthesize narratives. Lendol Calder (2006) recommends six core cognitive habits be taught to beginning students of history: questioning; contextualizing; sourcing; using evidence; recognizing multiple perspectives; and recognizing limits to what one knows. And Thomas Andrews and Flannery Burke (2007) put the power of alliteration

to work with “the Five Cs of Historical Thinking”: change, context, causality, contingency, and complexity.<sup>2</sup> The authors of these and other history primers concede that their chosen concepts and competencies reflect personal judgments and are neither comprehensive nor uncontroversial. In terms of reception, there seems to be little objection to specific items on the lists, just disagreements about the priority given to certain outcomes and questions about what level of undergraduate education they are best suited for. The many lists of core components of historical expertise define and prioritize the elements of historical thinking in different ways. But the most notable aspect of recent primers and handbooks is their unanimity on an important point: *that the main goal of history instruction should be historical thinking*. None frame student learning goals in terms of cultural literacy, historical knowledge, or specific content.

Pushing this conversation forward has been the work of cognitive scientists, particularly Sam Wineburg, on the mental processes that define historical expertise. As a doctoral student studying cognitive psychology at Stanford in the 1990s, Wineburg became intrigued by the elusive qualities of historical mindedness. Everything he read on this by historians, philosophers, and teachers came from introspective self-reporting, but he knew from research on expertise in other professions that a wide gap exists between what disciplinary experts say they do and what they actually do. Expert habits of mind are often invisible to the highly trained; they may forget there was a time when they did not know how to do what they have learned to do so well. To map the cognitive processes historians deploy, Wineburg used research protocols called think-alouds in which subjects are trained to think out loud while completing a task so that researchers can record and analyze their

---

<sup>2</sup> Allen Mikaelian, editor of the AHA's *Perspectives in History*, reports that Andrews and Burke's "Five Cs" essay is the most accessed article on the AHA Web archive. See *Perspectives on History* 52 (April 2014): 37.

introspections. Wineburg compared think-alouds conducted with historians to think-alouds with non-historians to learn how expert thinking differs from lay thinking. His empirical investigations showed that in history, experts and novices do not differ merely in *what* they know but, more crucially, in *how* they think. In particular, Wineburg stressed the ways expert historians source, corroborate, and contextualize evidence and the way they think about historical knowledge using concepts of significance, periodization, and narrative (Wineburg 2001).

Wineburg is but one of a number of cognitive psychologists who offer important insights into the mental operations of expert historians and clarify the often unstated expectations teachers hold for students. In a review of this literature in 2006, James Voss and Jennifer Wiley note that, compared with other fields, historians deal primarily with ill-structured problems that are not amenable to mathematics, formal logic, or repeatable experimentation. This leaves historians without common subject-matter knowledge and with more heterogeneity in the constraints they place on thinking. Nevertheless, like the authors of history primers, they found commonalities emerging in the literature. Voss and Wiley identified three interrelated tasks (historians obtain information, historians construct narratives, and historians make inferences and solve problems) and ten characteristics of history experts (CHEs). They report, for example, that as part of obtaining information, historians use sourcing heuristics to improve the structure of ill-structured problems, including corroboration, sourcing, contextualization, identifying absent evidence, and generating subtexts that illuminate the intentions of the author and thereby assist with the interpretation of sources. In constructing narratives, historians “provide rationales, explanations, elaborations, or speculations” in expository form and recognize the plausibility of alternative accounts (Voss and Wiley 2006, 575).

Ongoing work by cognitive scientists has made significant contributions to what we know about historical thinking and history

learning goals. Many of the CHEs found by Voss and Wiley echo lists found in historians' reflections on the discipline. Some of the differences are interesting, though. For example, they do not include recognizing contingency, a concept that many historians claim is crucial for explaining historical change. It remains to be seen through further research whether this is because historians do not actually work the concept the way we think we do or whether researchers have not adequately identified and mapped this cognitive move. In any case, cognitive scientists' precise, fine-grained unpacking of what it means to think like a historian is useful for framing and prioritizing learning outcomes in history. The influence of this literature on the development of history teaching materials and assessments is immense and growing.

In the context of standards-based reform and the so-called turn from teaching to learning, in the last two decades various parties have made notable efforts to engage more college faculty in discussions of learning outcomes for history students (Barr and Tagg 1995). Two of these efforts, the Quality in Undergraduate Education (QUE) project and the AHA's Tuning Project, show how much attitudes of history faculty members have changed in a short amount of time.

QUE, the first major national effort to articulate learning outcomes in selected disciplines, began in 1997 and ran until 2004. Sponsored by the Education Trust Inc., the National Association of Systems Heads, and Georgia State University, QUE funded faculty from twenty-one institutions in four states and five disciplines (including history) to meet twice a year in national workshops and periodically in local clusters. Using the lever of learning outcomes, QUE aimed to focus faculty attention on improving student learning by shifting attention away from what teachers say in lectures to what students were expected to know and do at various points in a history degree program. Organizers saw QUE as a professional development project that would disseminate pedagogical innovations like student learning outcomes, backwards course design,



rubrics for evaluating student work, and course-mapping tools for program improvement. However, the results disappointed expectations. Buy-in by faculty members and departments was mixed, with many history faculty members unwilling to concede that traditional teacher and content-centered pedagogies fail to generate higher level thinking. The learning outcomes produced by participating institutions varied greatly in quality; few saw the light of day as published models. The project that organizers and funders hoped would spark a revolution in higher education quietly died when its funding ended.

The story of the AHA Tuning Project, launched less than a decade later and funded by the Lumina Foundation, could not be more different.<sup>3</sup> Initially conceived to be a three-year project, it brings historians together to spell out the central skills, habits of mind, and understandings of the field of history in postsecondary education. Borrowing a model employed in European higher education, Tuning is a collaborative process in which participants define the core disciplinary elements of historical training and then harmonize, or tune, these goals in ways that are appropriate for their own institution's mission. Like QUE, Tuning does not aspire to standardize history curricula. Rather, it offers a process for departments to develop their own goals and curriculum while benefiting from conversation with others engaged in the same task. Unlike QUE, and critical for its success, Tuning has the AHA's authority behind it, making it a project led by historians for historians. Tuners have worked enthusiastically and productively in a collaborative process to create a Discipline Core Statement outlining what history students should know, do, and value. Whereas with QUE the identification of learning outcomes was merely the first step in a larger process of top-to-bottom reform of teaching and learning in the discipline, Tuning thus far restricts its scope to identifying learning outcomes. But in a crucial way Tuning is

---

<sup>3</sup> Coauthor Lendol Calder participated in both initiatives.

more ambitious than QUE, wanting to express history's core outcomes in ways that communicate the significance and value of a history degree to external audiences, including employers, policymakers, and students. Thus, departments engaged in Tuning are responsible for devising a "degree specifications profile" describing historical training at their institution and outlining core areas of competency expected of graduates from their program. The profiles released thus far draw from and adapt the Discipline Core statement, showing considerable overlap with concepts and competencies found in history primers (AHA, "About Tuning").

The surprisingly positive response to the original call for Tuners led the AHA to solicit applications for a second phase of the Tuning project. In January 2015, an even larger group of participants than the original sixty began working on their own degree specification profiles.

The success of Tuning shows how much has changed within the last decade. These changes likely reflect both developments within the profession—such as growing awareness of scholarship by historians on teaching and learning—and significant external pressures on history departments making them more receptive to defining learning outcomes with precision and transparency. Facing declining enrollments and pressure from administrators, accreditors, and policymakers to demonstrate the value of a history degree, more historians and departments have become engaged in conversations about defining and measuring student learning.

The AHA Tuning project indicates that there is considerable interest in the profession for the work of defining history learning outcomes. We believe the time may be right to carry this conversation to another level. It is one thing to theorize what it means to be good at history. It is another matter to consider how we might know whether students are achieving the goals we set for them. The interest and participation of over one hundred history departments in Tuning so far demonstrates strong interest among historians for thinking about what it means to be good at history

and how one becomes so, especially when that effort is led by historians. Now that Tuners have put substantial time and effort into defining history learning goals for their departments and worked to adapt their curriculums to achieve them, they will want to find meaningful ways to evaluate whether their efforts have improved student learning. This should make Tuners a receptive audience for tools that can help guide them in further program revisions. More than that, we expect Tuners to be participants and leaders in efforts to develop more authentic assessments of student learning in history.

### **Essential Concepts and Competencies for the History Major**

History primers, cognitive scientists, and the AHA Tuning Project have articulated comprehensive lists of student learning outcomes for history majors. We build on and advance this conversation by identifying a focused set of *essential concepts and competencies* that history faculty see as fundamental to the discipline, important enough to emphasize given limited time and resources, and valuable to students' lives. Focusing on a smaller number of outcomes enables more careful attention to how these core disciplinary goals can be learned, including more rigorous, targeted, and meaningful assessment. Moreover, framing these learning outcomes as *concepts* and *competencies* encourages historians to distinguish carefully between abilities and the conceptual understandings students must have to exercise those abilities.

In selecting these essential concepts and competencies, we looked for patterns of agreement in primers, cognitive science research, and current Tuning efforts. We relied on the input of the diverse historians in the MCL faculty panel, engaging with them in an iterative process of list making, feedback, winnowing, and further revision. Searching for areas of consensus in the field, we aimed to pinpoint *essential* outcomes, not a comprehensive list of

markers defining expertise in history. Consequently, our list likely leaves off concepts and competencies that individuals or departments deem important. In that event, they can and should articulate these additional outcomes and develop methods of assessment for them.

Perhaps the most obvious omission from our list is historical content knowledge. Clearly, historians value factual information about the past and consider subject matter literacy an important goal in their teaching. Furthermore, historical thinking requires content to function; historical concepts and competencies cannot be developed or practiced in a vacuum. The problem with including content knowledge as a goal for assessment is the question of *which* knowledge to test. Although histories of the United States and Europe once held privileged places in the curriculum of most colleges and universities, many departments no longer require immersion in these subjects. Instead, they encourage concentrations in other geographic areas and exploration of new thematic fields and faculty specializations. Thus, any attempt to build a test on a particular national history or to privilege particular regions or periods likely would meet with significant controversy. Furthermore, it is the conviction of many historians that no particular history *ought* to be privileged because historical thinking, the ultimate purpose of undergraduate history instruction, can be fostered in sustained study of any historical content. Within our MCL group we had passionate debates about content, specifically whether one's national history should have a privileged place. With no consensus possible and remembering that we were not attempting to be comprehensive, we decided to exclude any specific subject knowledge from the essential concepts and competencies.

### Concepts

Having reviewed what others before us have said in the ongoing conversation about history learning outcomes, we find the following

concepts are most essential for specifying what undergraduate history students should know at the completion of a course of study.

### **History as an Interpretative Account**

Students must understand that history is not simply what happened in the past unmediated by human sense-making. Rather, it is an interpretative account of the past constructed through a disciplined process of problem solving and supported by evidence that survives. Because we cannot apprehend the past through applications of mathematics, formal logic, or controlled experimentation, in historical accounts problem solving is usually verbal, with conclusions presented in the form of a narrative or an analytic argument developed in relation to particular questions, forms of evidence, and existing interpretations. Students must understand that when historians construct accounts their goal is not to reach a universal standard of validity or correctness as in the case of logical and mathematical proofs. Rather, the object is to convince an audience that an account of the past is highly acceptable. Evaluation of historical accounts occurs by examining the acceptability of the information provided as evidence, the extent to which the information supports the account, and the quality of counterarguments or alternative positions that may be offered. Both accounts and evaluations of accounts are influenced by historians' own beliefs, theoretical orientations, and other factors. Resting on interpretive accounts, it is the nature of historical knowledge to have relatively less certainty and more heterogeneity in how questions are answered than knowledge in some other disciplinary domains. This also means that historical knowledge is not fixed for all time. Rather, we can expect historical knowledge to be mutable. Historical knowledge is constantly being revised as new evidence comes to light and new generations ask different questions and attend to different constraints on our ability to know the past.

### The Relationship of Past and Present

Students must understand the complex relationship between past and present. Acceptable sense making of the past walks a balance between two states of mind: familiarity and strangeness. Often, we are motivated to study the past when we become aware that the world we live in today is a product of past events and developments that continue to shape contemporary life. This is the *presentness of the past*. But since nothing in time stands still, the passage of time makes strange what once seemed ordinary. Therefore, historians also emphasize the *pastness of the past*, that is, recognition of the differences that separate our own time from the past. Being mindful that the past is a foreign country cautions us to not assume we have an intuitive understanding of historical actors, projecting our own values and assumptions onto people of different times and places. Instead, recognizing the pastness of the past directs historians to understand people of the past by contextualizing their actions: what they were trying to accomplish; the nature of their beliefs, attitudes, and knowledge; the culturally and historically situated assumptions that guided thought and action. Situating people, events, and sources within the context of their time is a primary mission of historical sense making. Furthermore, examining the past gives a sense of the abnormal present; in other words, it helps to destabilize what we might take for granted in the present and help us view the present moment with critical perspective. Mediating between the pastness of the past and the presentness of the past gives people with a historical perspective a reflective self-awareness that actively searches for the plausibility of beliefs and actions different from their own.

### Historical Evidence

Students must understand that the acceptability of historical accounts depends a great deal on how evidence is used to support claims about the past. Sources of evidence are categorized as either

*primary* or *secondary*. Primary sources are the raw materials for the study of the past originating from the time under study. Secondary sources are interpretive accounts of the past that historians use to generate new questions, corroborate conclusions, and test interpretations. Students should understand that the classification of a source depends on its use for a particular historical question. For example, a historical account of ancient Rome written during the 1950s would be a secondary source if the historical question is about ancient Rome and the source is used to ground an interpretation, or it could be a primary source if one is asking a historical question about the intellectual climate affecting historians in the 1950s. Students should understand the nature, potential, and limits of both kinds of evidence. In particular, students should know that primary sources come in diverse forms, represent diverse perspectives, and have distinct strengths and limitations as evidence about the past. They will avoid the misconception that primary sources are exact, unproblematic reflections of the past. Critically, students must understand that reading primary sources for evidence demands a different approach than reading them for information. Acceptable interpretations of the past require that primary source evidence be examined both for content and its unwitting testimony, that is, what the source says without directly saying it. This requires asking questions about their provenance and historical contexts and using the answers to constrain interpretations of the evidence.

### **Complex Causality**

Students must understand that in contrast to disciplines that seek to isolate factors and reduce explanations to singular causes, history understands change over time to be complex and interconnected. Considerations include not only human agency but also structural, environmental, and other factors that play a role in stimulating, shaping, and resisting change. Thus, historical accounts are multiple

and layered, avoiding monocausal explanations and reductionist thinking. They distinguish significant from insignificant causes and proximate from long-term, enabling conditions. Causes put forward to explain an event (and the priority of causes) may differ based on the scale of the history and the approaches of the historian.

### **Significance**

Students must understand what makes something historically significant. Since the past is everything that happened before now, including everything that humans anywhere have thought, said, and done, no history can include all of the past. Therefore, the concept of significance is used to make choices about what subjects are worth remembering and constructing accounts about, what is worth including in an account, and what can be left out. Peter Seixas defines historical significance as “the valuing criterion through which the historian assesses which pieces of the entire possible corpus of the past can fit together into a meaningful and coherent story that is worthwhile” (Seixas 1994, 281). Historians generally regard something as significant if (a) it affects change or continuity with meaningful consequences, for many people, over a long period of time or if (b) it is revealing, leading us to understand other subjects in history and contemporary life in new ways, or was important at some stage in history within the collective memory of a group or groups.

### **Competencies**

Basing our judgment on a review of the history and literature on history learning outcomes, we recommend the following competencies as most essential for specifying what undergraduate history students should be able to do at the completion of a course of study.

#### **Evaluate Historical Accounts**

Students must recognize historical explanations in their most common forms: narrative, exposition, causal model, and analogy.



They should be able to identify an author's interpretation and critically scrutinize the evidence and analysis used to support it. In addition, they should be able to critically evaluate, compare, and synthesize historical accounts.

### **Interpret Primary Sources**

Students must be able to analyze and interpret information drawn from primary sources, drawing on specialized subject knowledge and concepts of historical thinking. More specifically, they should be able to distinguish primary from secondary sources; assess the credibility of sources and make judgments about their usefulness and limitations as evidence about the past; consider how the historical context in which information was originally created, accessed, and distributed affects its message; and address questions of genre, content, audience, perspective, and purpose to generate subtexts that illuminate the intentions of the author.

### **Apply Chronological Reasoning**

Students must take account of the role of time, sequencing, and periodization in historical narratives. In particular, students should demonstrate sensitivity to complex causation, with an ability to distinguish between proximate and ultimate causes; with a discerning eye for continuity and change over time; and with the ability to formulate and evaluate historical periods and turning points as heuristic devices for making sense of the past, recognizing the artificiality of periods and turning points and the ways they favor one narrative, theme, region, or group over others.

### **Contextualize**

Students must demonstrate the ability to place an event, actor, or primary source within the context of its time in order to interpret its meaning and significance. Rather than assume timeless, psychologized notions of why people behaved as they did in the past

or that people of the past were similar or identical to ourselves, with the same beliefs, attitudes, instincts, and motivations, students must be able to appreciate the particular policies, institutions, worldviews, and circumstances that shaped people's practices in a given moment in time. Recognizing difference is by itself not enough, however, if the past is dismissed for being unenlightened or immoral. Students must also be able to reconstruct the plausibility of other people's perspectives and actions within their own frame of reference. Contextualization does not mean identification (we can understand another's viewpoint without accepting it as our own), facile claims to knowing (we can never directly know others' experiences and perceptions), or an emotional response (the goal is understanding, not necessarily admiration or sympathy). Rather, students should be able to make sense of actions, social practices, and institutions in terms of people's reasons for doing or believing what they did.

### **Construct Acceptable Historical Accounts**

Students must be able to construct acceptable historical accounts that interpret the past using sources as evidence for knowledge claims in ways that demonstrate understanding of historical concepts, especially the nature of historical evidence, interpretation, and perspective. More specifically, students should be able to do the following: pose historical questions; select and utilize relevant and reliable primary source evidence to support their historical interpretation; extract information and supportable inferences from a wide range of primary and secondary sources, acknowledging, conceding, or refuting evidence that runs counter to the overall argument; recognize the limitations of evidence; and persevere through uncertainty, renouncing simple certitude (proof and inevitability) and easy relativism (every view is equal) for the disciplinary standard of limited relativism (plausible–implausible, acceptable–unacceptable).

### Reflections on the List of Essential Concepts and Competencies

A definitive list of learning outcomes for history is, of course, a chimera. We offer this list of essential concepts and competencies as our best summation of what history students in college should know, do, and value based on our study of a century and more of historians thinking out loud about history education and more recent attempts by historians, philosophers, history educators, and cognitive scientists to define the nature of expertise in history. Before we turn to the problem of how to design assessment tools that are worthy of historical understanding, we pause to consider two important questions raised by our list.

#### In college, what is the study of history for?

When thinking about learning outcomes for history education, questions about the purpose of a BA in history—or even the value of taking a single history course in college—cannot be avoided.

In today's career-minded environment, students are drawn to preprofessional programs because it is obvious what such majors train students to do. The premedical (premed) major prepares students for graduate education in health-related professions. The education major is for future teachers. The business major is for those who want to work in business fields. History programs suffer by comparison because a bachelor of arts in history is not a ticket to employment as a historian, and the demand for history MAs and PhDs is small. So why study history in college?

The answer, we believe, is that a course of study in history does two important things very well. In good history programs, the study of history effectively marries the analytic and synthesizing strengths associated with the liberal arts and sciences—America's premier educational tradition—with the problem-solving and practical strengths necessary to help companies and organizations succeed and grow. Better perhaps than most other disciplines or at least as well as any, history is positioned to help

people become civic and workplace leaders who think critically, communicate clearly, and solve complex problems. The reasons that this is so are apparent in our list of essential concepts and competencies.

In a paper published by the National History Center in 2008, “The Role of the History Major in Liberal Education,” Stanley Katz and James Grossman noted the close linkages between historical study and the broad aims of liberal learning. Katz and Grossman warn historians not to regard undergraduates as miniature graduate students, teaching history as the professors themselves were taught on the way to the PhD. Rather, they argue, undergraduate programs should be designed to “nurture [students’] liberal and civic capacities, in part by integrating disciplinary knowledge, methods, and principles into the broad experience of undergraduate education” (Katz and Grossman 2008). We concur wholeheartedly, and we call attention to the significant overlap between our list of essential history concepts and competencies and efforts by the Association of American Colleges and Universities (AAC&U) to invigorate liberal education for the 21st century through the decade-old LEAP initiative (Liberal Education and America’s Promise). To prepare students for responsible citizenship and a global economy, LEAP’s Essential Learning Outcomes for a liberal education are as follows: (a) knowledge of human cultures and the physical and natural world; (b) intellectual and practical skills (e.g., inquiry and analysis, critical and creative thinking, written and oral communication, and quantitative and information literacy); (c) teamwork and problem solving; (d) personal and social responsibility (e.g., civic knowledge and engagement—local and global, and intercultural competence); and (e) integrative and applied learning (e.g., synthesis across general and specialized studies demonstrated through complex problem solving). When we compare the LEAP outcomes with our list of essential concepts and competencies for history, we find that a course of study in history closely aligns

with the AAC&U's outline for liberal learning in the 21st century (AAC&U 2015).

But students may find another purpose for history education even more valuable, at least at first. Surveys of employers' priorities for the kinds of learning students need to succeed in today's competitive and global economy show that history is well positioned to provide what business and nonprofit leaders want. A 2013 AAC&U study conducted by Hart Research Associates found that 93 percent of employers believe "a demonstrated capacity to think critically, communicate clearly, and solve complex problems" is more important than a student's particular major. When employers were asked to endorse educational practices that would be helpful in preparing college students for workplace success, the practices they selected amount to a précis of the essential competencies we describe, including (a) conduct research and use evidence-based analysis; (b) gain in-depth knowledge and analytic, problem solving, and communication skills; and (c) apply learning in real-world settings (Hart Research Associates 2013). On the basis of surveys like this, we believe that history programs can capitalize by design on students' desire to prepare for career opportunities and success.

Why study history? An 1898 AHA pamphlet stated that "the chief purpose [of historical education] is not to fill the boy's head with a mass of material which he may perchance put forth again when a college examiner demands its production" (Sipress and Voelker 2011, 1051). By contrast, our list of concepts and competencies for a history BA foregrounds a very different purpose for undergraduate history education. The history BA prepares future civic and workforce leaders to grapple productively with ill-defined problems by bringing inquiry, analysis, and communication and application of knowledge to bear on specific complex questions. Historical study trains people to be citizens committed to liberal learning and innovative problem solvers in real-world settings.

**How do the essential concepts and competencies relate to introductory history courses?**

Unlike disciplines that offer one or two clearly delineated courses introducing students to the major, history has many pathways to the major and relatively little sequencing within it. Furthermore, introductory courses are taught in different ways and for different purposes. In some departments, introductory courses are small courses organized around doing history, aiming to introduce students to historical thinking and methodology through a focused topic. In many others—probably most—introductory courses remain broad surveys of historical knowledge and are commonly large lecture courses that also serve general education goals for the university. As our earlier summary of the history of learning outcomes suggests, in many of these courses the goal of cultural literacy and the methods of coverage are still embraced. Although some professors aim to infuse historical thinking into these surveys, many others emphasize broad exposure to content knowledge as the primary aim.

We believe that introductory courses can and should introduce students to the disciplinary concepts and competencies we propose and that this can exist alongside knowledge transmission goals. We believe this on the basis of the learning science that rejects the attic theory of cognition, which considers a stockpile of knowledge accumulated over years of study to be the prerequisite for advanced analytic work. We now know that students learn more when they are engaged early, often, and cumulatively in problem-centered inquiries requiring disciplined ways of thinking. Introductory courses, whether small seminars or large lectures, should aim to introduce and develop some or all of the essential concepts and competencies outlined here. However, we acknowledge that there might be considerable disagreement within the profession on this point, and some, perhaps many, historians view the traditional survey as serving legitimate, valuable, and important ends, including establishing a foundation of knowledge for later study or promoting

cultural and civic literacy. Many departments face real constraints that make doing history in small seminars seem unfeasible; reimagining the large introductory course to emphasize historical thinking, including its goals, methods, and assessments, is a project still in development.

Consequently, at this moment, even though we are confident that the learning outcomes we have defined are universally applicable to the BA in history, we do not believe the assessment imagined in this white paper will be considered appropriate by all instructors to measure student learning in introductory courses. Those still wedded to coverage methods as the best way to attain cultural literacy likely will object to our recommended assessment tools. On the other hand, those looking to build introductory courses on the platform of historical thinking for liberal learning and expanded opportunities in the workplace likely will be intrigued.

### **Student Learning in History: Past, Present, and Future Assessments**

Assessment is integral to history teaching and learning. History faculty members routinely assess student learning in individual courses, most often through papers and in-class examinations. Examinations in history often include multiple-choice questions, short answers (e.g., identifications), or essays that ask students to demonstrate knowledge or skills valued by the instructor. Essay assignments may ask students to analyze and synthesize a historical theme or historical narratives, interpret one or more primary sources, or conduct historical research by asking a historical question and answering it with primary and secondary sources. History faculty also employ a range of other assignments to assess student learning in courses: constructing primary source readers; oral presentations; short written assignments to assess particular skills like distinguishing between primary and secondary sources; questions for class discussion; and historical role-playing games like Reacting

to the Past.<sup>4</sup> Ideally, these assignments are carefully aligned with the instructor's learning goals, thereby providing valid evidence of students' achievement of those goals. Unfortunately, as Sipress and Voelker found in their analysis of introductory courses, though most historians claim to be teaching historical thinking, common assessments show otherwise. Constraints such as large class sizes and inadequate faculty development lead many instructors to settle for assessments requiring students to merely memorize and reproduce expert knowledge.

Less common are efforts to assess history learning across courses at the level of the history major. Most departments assume that successful completion of requirements, which may include standards for grade point average, breadth or specialization, and methods, demonstrates successful learning. However, an increasing number of history departments are making efforts to assess student learning more directly, using portfolios of student work, senior theses, capstone projects, or senior seminar courses to evaluate whether graduating seniors have "gotten it" and are able to demonstrate historical knowledge and thinking. Many of these efforts attempt to assess everything at once, providing relatively little feedback on which particular historical concepts and competencies majors have achieved. A thesis, for example, asks students to exhibit attainment of a great number of learning objectives simultaneously. Consequently, if the final product is unsatisfactory, in the absence of fine-grained rubrics it can be difficult to pinpoint the specific skills that need further development.

At the level of elementary and especially secondary education, there have been more efforts to develop large-scale assessments that provide information about student learning in history. The SAT Subject Tests offer exams in U.S. history and world history that can be taken by prospective college applicants and submitted

---

<sup>4</sup> See <https://reacting.barnard.edu> for more information on Reacting to the Past.



for college admission, although institutions rarely require them. The exams are composed entirely of multiple-choice questions that ask students to demonstrate breadth of knowledge above all else. The exams purport to incorporate social science concepts, methods, and generalizations by including short document excerpts including quotations and images as part of questions, but the overriding emphasis is on recall of content. A sample question in a practice guide, for example, asks students to identify the likely source of a quotation. The item requires examinees to read and comprehend the quotation, but ultimately it is assessing whether they can identify the individuals listed to discern the most likely author (College Board 2015).

The National Assessment of Educational Progress (NAEP) offers an assessment in U.S. history every five to eight years and is developing a test for world history. NAEP aims to measure what the nation's students in Grades 4, 8, and 12 know and can do in given subjects, rotating the subjects examined every year. NAEP exams use a matrix-sampling method in which every student is given only a portion of the test; rather than reporting the scores of individuals, NAEP reports the scores of groups of students. This allows them to examine students on a much wider range of materials. The U.S. history subject exam was first administered in 1986 as a multiple-choice exam that tested breadth of knowledge about U.S. history. In subsequent years and in light of the Bradley Commission, National History Standards, and burgeoning scholarship on history teaching and learning, NAEP decreased its multiple-choice portion to 50 percent, incorporated more open-ended responses and performance exercises on the exam, and infused ways of knowing and thinking about U.S. history into the four themes and eight periods that it examines. Today the NAEP U.S. history exam focuses strongly on breadth of historical knowledge, but it also uses multiple-choice questions and constructed responses to measure depth of learning and cognitive skills, including the ability of students to marshal facts and organize and express thoughts.

In some cases, students are given a textual or visual source and asked to respond to multiple-choice questions or write short open-ended responses to questions posed about it. The NAEP exam tests for historical knowledge and perspective and historical analysis and interpretation. The former asks students to identify, define, describe, and place knowledge, whereas the latter asks students to explain cause and effect, interpret different points of view, and define significance.

A third major history assessment in use today is perhaps the most well known to college faculty: the Advanced Placement exams in U.S. history, European history, and world history. Taken most often by high school juniors and seniors, these exams are meant to certify competence in the introductory college survey course to qualify for college credit. The exams are roughly half multiple-choice questions assessing students' knowledge of historical facts, one-quarter short essays measuring deeper knowledge about a topic and historical reasoning, and one-quarter a single document-based question (DBQ) that provides short primary source excerpts and asks students to write an essay interpreting the sources while drawing on background content knowledge. The examinations consequently involve writing and interpretation of sources but are heavily weighted toward assessing content knowledge. Recently the AP U.S. history exam underwent a redesign to address some of the criticisms of college faculty that it placed too much emphasis on recall of discrete facts, reinforcing a coverage model pedagogy and shortchanging important historical thinking skills. The new exam decreases multiple-choice questions to 40 percent of the score, emphasizes thematic learning objectives to help focus preparation for the exam and allow greater depth in teaching, and increases emphasis on historical thinking skills outlined in its curriculum framework.

These three major national assessments in history have several things in common and some important differences. All three examinations place strong emphasis on assessing specific content

knowledge about the past, be it U.S., European, or world history. Since the examinations are meant to assess learning in K–12 school subjects and, in the case of the AP exams, to substitute for a particular college course, this emphasis on demonstrating knowledge about major themes, events, and figures is understandable. NAEP and the AP exams have developed increasingly sophisticated frameworks for defining and assessing historical thinking within the context of a particular regional history; students might be asked, for example, to interpret a political cartoon that requires them not only to know something about the people and time being depicted but also to interpret a source. Yet for the reasons already outlined, an assessment of learning in the college history major will not place the same emphasis on evaluating a student's content knowledge. This raises thorny questions about assessment design that will be addressed in the next section: Since content knowledge is necessary for historical thinking, how can a history test be designed for diverse college curricula that require no common historical subject for the BA?

Another similarity between the three exams is that they all use multiple-choice questions to assess student knowledge and, to some extent, historical cognition. For large-scale standardized assessments, reliance on multiple-choice questions is unsurprising: They are easy, fast, and inexpensive to administer and score. In a shorter amount of time and at less cost than other methods of assessment, multiple-choice items can gather more information about student learning. Assessment experts also have great faith in the validity (how well they represent the learning outcome desired) and reliability (how consistent the results are) of multiple-choice items once they have been rigorously field tested and refined.

However, we caution that multiple-choice items might face a hostile reception among college history faculty. Some researchers in history teaching and learning have rejected multiple-choice questions as appropriate assessments for the field because they do not reveal the cognitive processes behind an answer.

Moreover, multiple-choice items privilege a single, right answer when historical interpretation is by nature complex, nuanced, and multiple, allowing for several acceptable answers. Some have raised concerns that multiple-choice items can be skewed by strong test-taking skills; students can puzzle out the correct response even when they cannot come up with it on their own (Ercikan and Seixas 2015). Discussions with our MCL faculty panel indicated that most historians in the room associated multiple-choice items with memorization of facts and lower order thinking skills. Even when given questions from the redesigned AP U.S. history exam that aim to measure deeper cognitive processes, the MCL historians remained skeptical, faulting the validity of the items. When items are subject to multiple interpretations, will the best students choose the wrong answers? Some historians in the group did!

A final observation about existing large-scale assessments is that the most respected tests, NAEP and AP, use constructed responses—open-ended writing—to evaluate knowledge and thinking in deeper ways. Short constructed responses and essay questions ask students to demonstrate knowledge, make and support claims about changes and continuity over time, identify relevant and significant information, and organize and cogently communicate ideas. Consequently these exams allow more opportunity to gather evidence about students' cognitive moves, including their mastery of concepts and competencies. The DBQ in particular has often been regarded as a gold standard in history assessment because it is an authentic assessment, meaning it assesses students' ability to complete a task that replicates the work historians actually do. The faculty panel was relatively impressed by the DBQ, although some noted ways it might be adjusted to measure additional facets of historical thinking. For example, some historians thought the document excerpts were too obviously targeted to the question, a problem that could be fixed with longer excerpts and the inclusion of less useful or relevant sources to evaluate students' abilities to evaluate and select sources. Others noted potential for

a DBQ to include secondary sources or to be designed around evaluating historical interpretations instead of only primary sources.

The DBQ is highly regarded, but a difficulty with this type of assessment is that it measures many different learning outcomes simultaneously, meaning the information it gives about student learning is not as clear and precise as more targeted assessments that isolate concepts and competencies. This weakness has inspired the Stanford History Education Group (SHEG), headed by Sam Wineburg, to develop and make available to teachers through its free website a range of short constructed-response items called History Assessments of Thinking (HATs) that can be used in the classroom to inform instruction. HATs isolate and measure specific cognitive moves that define historical thinking, such as sourcing, corroborating, and contextualizing. For example, one HAT gives students the 1932 painting *The First Thanksgiving 1621* depicting Native Americans and Pilgrim settlers sharing a meal. The title, painter, and date are given along with the prompt, “The painting *The First Thanksgiving 1621* helps historians understand the relationship between the Wampanoag Indians and the Pilgrim settlers in 1621. Do you agree or disagree? (Circle one.) Briefly support your answer.” Students are given several lines to answer, which is enough space to explain their reasoning but requires them to be succinct. This particular HAT evaluates students’ ability to source material by evaluating whether they understand the importance of the painting date—1931—in relation to the event being depicted. HATs are designed to be relatively fast assessments to administer and score so that teachers can use them in the classroom as formative, or baseline, assessments. The benefit of this approach is twofold. First, unlike a complex task like the DBQ, HATs allow assessors to isolate specific concepts and competencies, thereby supplying more targeted information about student learning. Second, unlike multiple-choice exams, HATs do not assume a single, right answer but instead put the focus on how students justify their conclusions. In the *First Thanksgiving* example, one

can imagine strong responses that both agree and disagree with the prompt. One might disagree that the painting helps historians understand the relationship between natives and settlers because it was painted over three hundred years later and was based on the artists' imagination, not historical research. However, one might also agree with the statement by stipulating reasons to believe the painter based the depiction on historical documents and aimed for historical accuracy. HATs provide an innovative model for assessing historical thinking that might be expanded to target additional learning outcomes, including specific skills with secondary texts as well as primary sources (SHEG, "Beyond the Bubble").

It is also possible to imagine new forms of assessment that utilize technology to create authentic assessments that mimic the work historians do. In other disciplines, testing companies are experimenting with computer simulations, such as performing a virtual scientific experiment or cooperating on a group problem-solving task. These interactive assessments allow students to perform a task that measures essential concepts and competencies in that discipline. In history, one might imagine a virtual archive, for example, that allows students to find, select, and utilize sources out of many possibilities, assessing their ability to ask a historical question, sift and identify relevant and significant sources, and interpret those sources. One might also imagine a simulation that asks students to role play in some fashion, demonstrating ability to take perspectives, understand context, and reason historically.

### **An Assessment for the History Major: Form, Uses, and Next Steps**

Historians should use multiple forms of assessment to gather as much evidence as possible about student learning in history. Different assessments can focus on evaluating different things and for different purposes: for example, assessment can focus on certifying the learning of individual students or it can be used to evaluate a

department's curriculum by focusing on the performance of majors as a whole; it can be used to inform and improve instruction as it is happening or to measure achievement of learning outcomes as seniors finish the major. Multiple forms of assessment allow stronger conclusions about student learning and ensure that no single measure is invested with so much authority that it narrows or distorts teaching and learning. Historians already have some assessments available—like portfolios and senior theses—and we urge that departments think about how to utilize these in more systematic ways to gather evidence about student learning that will inform teaching and curriculum. In addition, however, we believe that rigorously developed, standardized instruments can be a valuable contribution to these efforts and have important uses for history departments.

As a first step, we recommend the development of a voluntary, standardized assessment under the control of departments to evaluate whether students completing their history major have mastered foundational disciplinary concepts and competencies in history. We hope this is the first of many such tools developed for departments and that standardized instruments form only one component of a larger assessment toolkit.

A professionally developed assessment of the major at the level of the department has several potential benefits. In the first place, it will allow departments to collect and analyze evidence about the achievement of graduating seniors. In doing so, it can help departments to identify strengths and weaknesses in student learning that can guide curriculum development and instructional choices. In addition, this evidence can help strengthen the case to students, parents, university administrators, employers, and the larger public about the value of the history major. The AHA Tuning Project has taken an important first step in making this case but a trustworthy, objective assessment of student learning would strengthen it considerably. Relatedly, many departments face pressure from university administrators or external entities likes

accreditors, state legislatures, and regulatory bodies to demonstrate their effectiveness. An assessment of student learning in the major provides far more meaningful evidence of departments' work than most other measures available. Perhaps it can also direct conversations about assessment that are already gaining steam in recent years into directions that history faculty find more valuable and legitimate. In states like Texas, for example, a growing conversation about assessment has many faculty fearing, not without reason, that externally imposed assessments will draw from models like the SAT Subject Tests that prioritize low-order thinking skills and memorization of historical facts, thereby narrowing and misdirecting history teaching and learning. In developing an assessment that has legitimacy among college history faculty, perhaps we can help to shape the conversation and the efforts that are bound to continue in the coming years. Finally, a standardized assessment has the potential to offer new knowledge about history teaching and learning. It could provide a large body of evidence about student learning in history that will allow researchers to ask and answer important questions.

In thinking about the forms that an assessment for the major might take, it is essential that the discussion continues among history faculty and that historians take the lead role in defining the goals, forms, and uses of assessment. With that in mind, we urge consideration of a few key issues and offer some preliminary recommendations to serve as a starting point for discussions.

First, careful attention should be given to the issue of the *assessment's primary purpose*. Specifically, should the test measure and report individual student scores, or should it instead measure and report at the aggregate level of the department? The assessment could be designed in a way to report individual students' scores, enabling tracking of specific students' learning over time if the assessment is administered at different points, such as at the time of declaring the major and at the time of graduation. Aggregated, these scores could serve as a measure of the effectiveness of the



curriculum overall. But individual reporting could have other uses, too. Students might want to use the test for signaling purposes when applying for jobs. Instructors could use individual scores as formative assessment for guidance purposes; it could help to diagnose areas for improvement and additional emphasis to guide students in their history studies. Alternately, the assessment could eschew reporting individual student scores and instead report about student learning as a group as NAEP and the Collegiate Learning Assessment (CLA) do, effectively making the department as a whole the object of assessment. One benefit of this approach is that it allows for matrix sampling, meaning that students can be given different versions of the assessment and therefore evaluated on a much broader range of items and in less time. If the object is to measure departments rather than individuals, it also gives more flexibility for test items including performance tasks since the standards for validity for measuring group performance versus individuals are different.

Second, it is important that a history assessment have substantial focus on revealing cognitive processes of historical thinking, not simply assessing the ability to arrive at a correct answer. Short constructed-response items like HATs may be the best way to do this. However, it is possible that these cognitive processes can be assessed using carefully constructed multiple-choice questions, an approach that has important benefits from the standpoint of assessment: namely, multiple-choice items are easier and less costly to score and faster to take, allowing the assessment to collect more discrete pieces of information about student learning in a faster amount of time and for less cost. Recently, Bruce VanSledright has put forward suggestions for how multiple-choice tests might be made more responsive to the complexity of historical thinking by providing statements that are weighted according to their defensibility (VanSledright 2013). If multiple-choice approaches are pursued, however, it is essential to first explore their legitimacy among historians, because if our MCL panel is representative, there may be too much skepticism among academic historians to use them.

Third, the assessment should have a strong authentic component, meaning it asks students to demonstrate the concepts and competencies in performance tasks that approximate the real work historians do, namely, analyzing sources and constructing historical explanations and arguments in writing. An assessment of history majors that does not have a significant writing component or ask students to do history in some form will not be accepted as legitimate by many college faculty. Furthermore, although assessing concepts and competencies individually offers important information about student learning, it is also important that students synthesize them because the concepts and competencies are inter-related. Consequently, a DBQ-type performance exercise that asks students to demonstrate the ability to think historically should be an important component of the assessment.

Finally, a significant issue to address in designing a history assessment is the question of which subject matter content will be used in the assessment. For reasons outlined earlier, historical knowledge is not among the essential concepts and competencies that would be assessed. Theoretically, then, the content of the exam should not matter since the assessment will be rigorously field tested and validated to ensure that it measures a given concept or competency regardless of a student's background knowledge. However, we believe the choice of content for the exam will matter greatly to historians, both in terms of what it communicates about how the test defines and privileges particular fields, and because many of us suspect that no matter how rigorously developed, a history assessment cannot and ought not to entirely strip historical content away. Students will bring content knowledge to the exam that will shape their encounter with test items. Content informs and is embedded in most facets of our core disciplinary concepts and competencies: how does one demonstrate proficiency in contextualization without being able to draw on knowledge of a particular time and place to locate a source or event in context? Consequently the question of what content will be on the exam is a

thorny, but important, issue to resolve. Existing history assessments do not face this issue since they assess historical knowledge and ground historical thinking skills within a specified body of content knowledge, be it U.S., European, or world history.

Based on our discussions with the MCL faculty panel, assessment experts, and one another, we offer the following description of a potential assessment as a starting point for further discussion among history faculty. We propose that the test contain two distinct parts. The first part would assess historical concepts and competencies in isolation or in small combinations to give targeted information on student learning. We recommend that this part be composed entirely of HAT-type short constructed responses or, if further discussions in the field warrant it, half HAT-type responses and half multiple-choice questions. This part of the exam could draw its questions from all regions and time periods and would have to be rigorously field tested to ensure that it assessed the given historical concept or competency rather than content knowledge. Students might have to evaluate historical accounts from early modern Britain, source primary texts from 19th-century Ghana, and interpret primary sources from 20th-century Mexico. It may be that the assessment could supply necessary background knowledge to isolate the concept or competency. To keep the time needed to take this section reasonable, it may be that a single piece of evidence could be used to ask several distinct questions or that these HATs could be combined with the sources used in the second part of the exam.

The second part of the assessment would be a DBQ or similar lengthy performance task that asks students to do history in a meaningful way, drawing on most concepts and competencies in the process. For part two, students should be given a choice of questions that would allow them to select a DBQ for a geographic area or time period in which they have some familiarity through course work. A computer-based assessment would make this particularly feasible.

The exam should be administered at point of entry to the major and again at completion to allow departments to assess learning within the major over time. We recommend that scores be reported to institutions and under their control (like the CLA is currently done) to avoid sensationalized comparisons and allay some of the concerns of historians about how external authorities might misuse these assessments.

In moving forward there are two major problems to address: (a) the technical problems and cost of developing an assessment; and (b) suspicion and mistrust of college faculty to standardized assessment. The technical challenges and cost can be overcome but will require substantial interest and investment by one or more organizations with expertise in assessment such as the Educational Testing Service, the Stanford History Education Group, or other entities. Proposed items must be rigorously field tested and validated against other kinds of measures to ensure that each task or question measures what it claims to measure and does so without measuring other things that will skew the results and weaken the inferences that can be made about student learning. These problems could only be discovered through repeated trials and adjustment of items and by comparing results with alternative forms of assessment meant to measure the same learning goals. Developing a good assessment is far more difficult, involved, and technical than simply having an expert produce a set of tasks she believes is good. Given the time and expense involved, for historians to undertake development of an assessment tool, we would have to be confident about finding a receptive market.

This leads to the second challenge: although there are encouraging signs that some historians and departments are receptive to framing and assessing learning outcomes, on the whole, there is still considerable suspicion, even outright opposition, among college history faculty to the project of assessment. This opposition is rooted for some in philosophical objections—rejection of the very project of defining behavioral objectives and attempting to

measure them—and for many others on the experience of testing at the K–12 level, including the lower order fact retention many of us associate with standardized history assessment and the perverse effects that testing has had on K–12 education in general. We share these concerns.

Yet the inability to measure everything we value about history—such as its ability to stimulate curiosity and generate meaning—does not mean historians can abdicate our responsibility to think carefully about what we *can* assess and use it to improve our teaching. Furthermore, our critiques of K–12 assessment examples should inspire us to engage in and shape the conversation, not run from it. Historians can choose to take a leading role in designing assessments that are worthy of the history we teach, or we can do nothing and hope that calls for accountability in higher education will go away. The latter course of inaction, we fear, will see historians being forced to assess courses and programs with odious tests designed by others.

Opposition within the field is a serious obstacle since a disciplinary assessment meant to measure student learning in history at the level of the department will be effective and meaningful only if faculty accept it as legitimate and use it to inform and improve practice. The experience of QUE suggests such a test will fail in its goals if it is imposed by institutions or external authorities on unwilling departments. Hence, the role of the American Historical Association will be key. As the chief organization of professional historians in the United States, it has an authority and legitimacy among college history faculty that cannot be matched by any other group. AHA has already taken a leadership role in conversations about outcomes. We urge the AHA to stay in the game and continue to lay the groundwork for further conversations about assessment and pedagogies, supported by foundation funding.

One obvious place for AHA to encourage this work would be to expand the scope of the Tuning Project. Tuners might be a particularly receptive audience for conversations about assessment

since they already have conducted lengthy discussions about departmental goals. The obvious next question concerns how to know whether these goals are being met. Volunteers could be solicited from departments who have already completed the Tuning process to take the next step and begin planning for assessment of the learning outcomes they have defined. As in Tuning, the process itself will be as important as the product. The search for department-level assessment should model Tuning's emphasis on deliberation and its careful balance of external support and guidance with respect for institutional autonomy and contexts.

Another receptive audience would be historians in states that are already moving to require standardized assessment. These conversations might begin with some of the issues raised in this white paper, including the refinement of essential concepts and competencies, the best models for assessment, the legitimacy of multiple-choice questions, and the question of how to deal with content on the exam. The goal would be to take the temperature of the profession on these issues and work toward areas of consensus and to foster individuals' and departments' engagement with the issues, which we believe is the only way to gain buy-in and support for assessment.

AHA might spearhead conversations about assessment within history in other ways that would be most effective if supported by a foundation or other fiscal support. It might, for example, organize panels at the AHA annual meeting or hold special smaller meetings or workshops devoted to the subject. It might offer seed grants to a small number of departments (perhaps former Tuners) to develop institution-specific plans for assessment or to work on recommendations for a standardized assessment for the field. AHA might create a special section of its website where it collects and shares information on assessment for history departments who are working to develop their own plans. The website could collect useful links, such as to the assessment resources of the Stanford History Education Group, and could provide concrete examples of

what institutions are doing to assess student learning among history majors, such as a description of how history faculty at California State University, Long Beach use and evaluate portfolios of student work. Even if a standardized assessment for the discipline is never developed, it is important that departments work to be more deliberate, systematic, and thoughtful about student learning in the major and how to evaluate it. AHA can perform a real service by helping to organize discussions of assessment and share best practices among institutions so that individual departments are not required to reinvent the wheel.

We have argued that a standardized history assessment would be a valuable tool for capturing usable information about student learning. One reason for doing so is to be ready with answers when students, parents, administrators, and public representatives ask about the value of a history degree. Of course, departments can hand out degree specifications and other documents articulating how the study of history instills crucial skills and habits of mind for citizenship, employment, and personal well-being. And history teachers can point to their grade books as evidence of achievement in these areas. But is this really the best we can do, asking the public to take our word for it? In this moment of crisis in higher education, historians need an assessment tool worthy of our discipline, one that will measure the most crucial concepts and competencies gained in the course of undergraduate study in history. We need such a tool not only to justify the value of our degree programs to others but also for ourselves. As historians, we know that corroborated evidence is preferable to no evidence or just a little evidence. The information gained from a standardized history assessment tool could be used to strengthen history curricula at the department level and inform research about history teaching and learning. For this, historians do not need a crisis; already we have our curiosity and responsibility as professionals.

Thus, we urge our fellow historians to carefully consider the potential—and the risks—of new assessment tools for history

teaching and learning, especially those that would inform decisions about improving the history major as a whole. It is not enough to make stirring pronouncements about what a history degree aims to teach. We also should attend to evidence for what students are learning.

## References

- Association of American Colleges and Universities (AAC&U). 2015. *The LEAP Challenge: Education for a World of Unscripted Problems*. Washington, DC: AAC&U.
- American Historical Association (AHA). 2013. "AHA History Tuning Project: History Disciplinary Core." <http://www.historians.org/teaching-and-learning/current-projects/tuning/history-discipline-core>.
- American Historical Association (AHA). n.d. "About Tuning." <http://www.historians.org/teaching-and-learning/current-projects/tuning/about-tuning>.
- Andrews, Thomas, and Flannery Burke. 2007. "What Does It Mean to Think Historically?" *Perspectives on History* 45: 32–35.
- Barr, Robert B., and John Tagg. 1995. "From Teaching to Learning: A New Paradigm for Undergraduate Education." *Change* 27, no. 6: 13–26.
- Bloch, Marc. 1953. *The Historian's Craft*. New York: Alfred Knopf.
- Bradley Commission on History in Schools. 1989. "Building a History Curriculum: Guidelines for Teaching History in Schools." *The History Teacher* 23, no. 1: 7–35.
- Calder, Lendol. 2006. "Uncoverage: Toward a Signature Pedagogy for the History Survey." *Journal of American History* 92: 1358–1369.
- Calder, Lendol, William W. Cutler III, and T. Mills Kelly. 2002. "History Lessons: Historians and the Scholarship of Teaching and Learning." *Disciplinary Styles in the Scholarship of Teaching and Learning*, edited by Mary Taylor Huber and Sherwyn P. Morreale, 45–68. Washington, DC: American Association for Higher Education and The Carnegie Foundation for the Advancement of Teaching.



- Carr, Edward H. 1961. *What Is History?* New York: Cambridge University Press.
- College Board. 2015. *Getting Ready for the SAT Subject Tests*. New York: College Board.
- Collingwood, R. G. 1939. *An Autobiography*. New York: Oxford University Press.
- Collingwood, R. G. 1946. *The Idea of History*. New York: Oxford University Press.
- Ercikan, Kadriye, and Peter Seixas, Eds. 2015. *New Directions in Assessing Historical Thinking*. New York: Routledge.
- Evans, Richard J. 2000. *In Defense of History*. New York: W.W. Norton.
- Flaherty, Colleen. 2015. "Text Anxiety." *Inside Higher Education*, January 28. <https://www.insidehighered.com/news/2015/01/28/purdue-president-and-faculty-clash-over-student-learning-assessment>.
- Florida K-20 Education Code. 2015. "Public K–12 Educational Instruction: Required Instruction." 1003.42.
- Furay, Conal, and Michael J. Salevouris. 1988. *The Methods and Skills of History: A Practical Guide*. Wheeling, IL: Harlan Davison.
- Gaddis, John Lewis. 2004. *The Landscape of History: How Historians Map the Past*. New York: Oxford University Press.
- Hart Research Associates. 2013. *It Takes More than a Major: Employer Priorities for College Learning and Student Success: Overview and Key Findings*. Washington, DC: AAC&U.
- Holt, Thomas. 1995. *Thinking Historically: Narrative, Imagination, and Understanding*. New York: College Entrance Examination Board.
- Jenkins, Keith. 1991. *Re-Thinking History*. New York: Routledge.
- Katz, Stanley N., and James R. Grossman. 2008. *The History Major and Undergraduate Liberal Education: Report of the National History Center Working Group to the Teagle Foundation*. Washington, DC: American Historical Association.
- Lerner, Adam. 2015. "History Class Becomes a Debate on America." *Politico*, February 21. <http://www.politico.com/story/2015/02/ap-us-history-controversy-becomes-a-debate-on-america-115381.html>.
- National Commission on Excellence in Education. 1983. *A Nation at Risk: The Imperative for Educational Reform*. Washington, DC: U.S. Department of Education.

- Pace, David. 2004. "The Amateur in the Operating Room: History and the Scholarship of Teaching and Learning." *American Historical Review* 109: 1171–1192.
- Rothstein, Richard. 2004. "We Are Not Ready to Assess History Performance." *Journal of American History* 90: 1381–1391.
- Seixas, Peter. 1994. "Students' Understanding of Historical Significance." *Theory and Research in Social Education* 22, no. 3: 281–304.
- Sipress, Joel M., and David Voelker. 2011. "The End of the History Survey Course: The Rise and Fall of the Coverage Model." *Journal of American History* 97: 1050–1066.
- Stanford History Education Group (SHEG). n.d. "Beyond the Bubble: A New Generation of History Assessments." <https://beyondthebubble.stanford.edu/assessments>.
- Taylor, Paul, Kim Parker, Richard Fry, D'Vera Cohn, Wendy Wang, Gabriel Velasco, and Daniel Dockterman. 2011. *Is College Worth It? College Presidents, Public Assess, Value, Quality, and Mission of Higher Education*. Report of the Pew Research Center. Washington, DC: Pew Social & Demographic Trends.
- VanSledright, Bruce A. 2013. *Assessing Historical Thinking and Understanding: Innovative Designs for New Standards*. New York: Routledge.
- Voss, James F., and Jennifer Wiley. 2006. "Expertise in History." In *The Cambridge History of Expertise and Expert Performance*, edited by K. Anders Ericsson, Neil Charness, Robert R. Hoffman, and Paul J. Feltovich, 569–586. New York: Cambridge University Press.
- Wineburg, Sam. 2001. *Thinking Historically and Other Unnatural Acts: Charting the Future of Teaching the Past*. Philadelphia: Temple University Press.