

Sam Wineburg. "Crazy for History." *The Journal of American History*, Vol. 90, No. 4, March 2004.

In 1917, the year the United States went to war, history erupted onto the pages of the American Psychological Association's *Journal of Educational Psychology*. J. Carleton Bell, the journal's managing editor and a professor at the Brooklyn Training School for Teachers, began his tenure with an editorial entitled "The Historic Sense." (A companion editorial examined the relation of psychology to military problems.) Bell claimed that the study of history provided an opportunity for thinking and reflection, qualities lacking in many classrooms.¹

Bell invited his readers to ponder two questions: "What is the historic sense?" and "How can it be developed?" Such questions, he asserted, did not concern only the history teacher; they were ones "in which the educational psychologist is interested, and which it is incumbent upon him to attempt to answer." To readers who wondered where to locate the elusive "historic sense," Bell offered clues. Presented with a set of primary documents, one student produces a coherent account while another assembles "a hodgepodge of miscellaneous facts." Similarly, some college freshmen "show great skill in the orderly arrangement of their historical data" while others "take all statements with equal emphasis . . . and become hopelessly confused in the multiplicity of details." Did such findings reflect "native differences in historic ability" or were they the "effects of specific courses of training"? Such questions opened "a fascinating field for investigation" for the educational psychologist.²

Bell's questions still nag us today. What is the essence of historical understanding? How can historical interpretation and analysis be taught? What is the role of instruction in improving students' ability to think? In light of his foresight, it is instructive to examine how Bell carried out his research agenda. In a companion article to his editorial, Bell and his colleague David F. McCollum presented a study that began by laying out five aspects of the historic sense:

1. "The ability to understand present events in light of the past."
2. The ability to sift through the documentary record—newspaper articles, hearsay, partisan attacks, contemporary accounts—and construct "from this confused tangle a straightforward and probable account" of what happened.
3. The ability to appreciate a historical narrative.
4. "Reflective and discriminating replies to 'thought questions' on a given historical situation."
5. The ability to answer factual questions about historical personalities and events.³

The authors conceded that the fifth aspect was "the narrowest, and in the estimation of some writers, the least important type of historical ability." Yet, they acknowledged, it was the "most readily tested." In a fateful move, Bell and McCollum elected the path of least resistance: of their five possibilities only one—the ability to answer factual questions—was chosen for study. While perhaps the first instance, this was not the last in which ease of measurement—not priority of subject matter understanding—determined the shape and contour of a research program.⁴

Bell and McCollum created the first large-scale test of factual knowledge in United States history and administered it to fifteen hundred Texas students in 1915–1916. They compiled a list of names (for example, Thomas Jefferson, John Burgoyne, Alexander Hamilton, Cyrus H. McCormick), dates (1492, 1776, 1861), and events (the Sherman Antitrust Act, the Fugitive Slave Act, the Dred Scott decision) that history teachers said every student should know. They administered their test at the upper elementary level (fifth through seventh grades), in high schools (in five Texas districts: Houston, Huntsville, Brenham, San Marcos, and Austin), and in colleges (at the University of Texas, Austin, and at two teacher-training institutions, South-West Texas State Normal School and Sam Houston Normal Institute).

Across the board, results disappointed. Students recognized 1492 but not 1776; they identified Thomas Jefferson but often confused him with Jefferson Davis; they uprooted the Articles of Confederation from the eighteenth century and plunked them down in the Confederacy; and they stared quizzically at 1846, the beginning of the U.S.-Mexico war, unaware of its place in Texas history. Nearly all students recognized Sam Houston as the father of the Texas republic but had him marching triumphantly into Mexico City, not vanquishing Antonio Lopez de Santa Anna at San Jacinto.

The overall score at the elementary level was a dismal 16 percent. In high school, after a year of history instruction, students scored a shabby 33 percent, and in college, after a third exposure to history, scores barely approached the halfway mark (49 percent). The authors concluded that studying history in school led only to "a small, irregular increase in the scores with increasing academic age." Anticipating jeremiads by secretaries of education and op-ed columnists a half century later, Bell and McCollum indicted the educational system and its charges: "Surely a grade of 33 in 100 on the simplest and most obvious facts of American history is not a record in which any high school can take great pride."⁵

By the next world war, hand-wringing about students' historical benightedness had moved from the back pages of the *Journal of Educational Psychology* to the front pages of the *New York Times*. "Ignorance of U.S. History Shown by College Freshmen," trumpeted the headline on April 4, 1943, a day when the main story reported that George Patton's troops had overrun those of Erwin Rommel at Al-Guettar. Providing support for the earlier claim made by the historian Allan Nevins that "young people are all too ignorant of American history," the survey showed that a scant 6 percent of the seven thousand college freshmen tested could identify the thirteen original colonies, while only 15 percent could place William McKinley as president during the Spanish- American War. Less than a quarter could name two contributions made by either Abraham Lincoln or Thomas Jefferson. Often, students were simply confused. Abraham Lincoln "emaciated the slaves" and, as first president, was father of the Constitution. One graduate of an eastern high school, responding to a question about the system of checks and balances, claimed that Congress "has the right to veto bills that the President wishes to be passed." According to students, the United States expanded territorially by purchasing Alaska from the Dutch, the Philippines from Great Britain, Louisiana from Sweden, and Hawaii from Norway. A *Times* editorial excoriated those "appallingly ignorant" youth.⁶

The *Times's* breast-beating resumed in time for the bicentennial celebration, when the newspaper commissioned a second test, with Bernard Bailyn of Harvard University leading the charge. With the aid of the Educational Testing Service (ETS), the *Times* surveyed nearly two thousand freshmen on 194 college campuses. On May 2, 1976, the results rained down on the bicentennial parade: "*Times* Test Shows Knowledge of American History Limited." Of the 42 multiple-choice questions on the test, students averaged an embarrassing 21 correct—a failing score of 50 percent. The low point for Bailyn was that more students believed that the Puritans guaranteed religious freedom (36 percent) than understood religious toleration as the result of rival denominations seeking to cancel out each others' advantage (34 percent). This "absolutely shocking" response rendered the voluble Bailyn speechless: "I don't know how to explain it."⁷

Results from the 1987, 1994, and 2001 administrations of the National Assessment of Educational Progress (NAEP, known informally as the "Nation's Report Card") have shown little deviation from earlier trends.⁸ In the wake of the 2001 test came the same stale headlines ("Kids Get 'Abysmal' Grade in History: High School Seniors Don't Know Basics," *USA Today*); the same refrains of cultural decline ("a nation of historical nitwits," wagged the *Greensboro* [North Carolina] *News and Record*); the same holier-than-thou indictments of today's youth ("dumb as rocks," hissed the *Weekly Standard*); and the same boy-who-cried-wolf predictions of impending doom ("when the United

States is at war and under terrorist threat," young people's lack of knowledge is particularly dangerous).⁹ Scores on the 2001 test, after a decade of the "standards movement," were virtually identical to their predecessors. Six in ten seniors "lack even a basic knowledge of American history," wrote the *Washington Post*, results that NAEP officials castigated as "awful," "unacceptable," and "abysmal." "The questions that stumped so many students," lamented Secretary of Education Rod Paige, "involve the most fundamental concepts of our democracy, our growth as a nation, and our role in the world." As for the efficacy of standards in the states that adopted them, the test yielded no differences between students of teachers who reported adhering to standards and those who did not. Remarked a befuddled Paige, "I don't have any explanation for that at all."¹⁰

To many commentators, what is at stake goes beyond whether today's teens can distinguish whether eastern bankers or western ranchers supported the gold standard.¹¹ Pointing to the latest NAEP results, the Albert Shanker Institute, sponsored by the American Federation of Teachers, claimed in a blue-ribbon report, "Education for Democracy," that "something has gone awry.... We *now* have convincing evidence that our students are woefully ignorant of who we are as Americans," indifferent to "the common good," and "disconnected from American history."¹²

One wonders what evidence this committee "now" possesses that has not been gathering moss since 1917 when Bell and McCollum hand tallied fifteen hundred student surveys. Explanations of today's low scores disintegrate when applied to results from 1917—history's apex as a subject in the school curriculum.¹³ No one can accuse the Texas teachers of 1917 of teaching process over content or serving up a tepid social studies curriculum to bored students—the National Council for the Social Studies (founded in 1921) did not even exist. Instead of being poorly trained and laboring under harsh conditions with scant public support, the Texas pedagogues were among the most educated members of their communities and commanded wide respect. ("The high schools of Houston and Austin have the reputation of being very well administered and of having an exceptionally high grade of teachers," wrote Bell and McCollum—it is hard to imagine that sentence being written about today's schools.)¹⁴

Historical memory shows an especial plasticity when it turns to assessing young people's character and capability. The same Diane Ravitch, educational historian and member of the NAEP governing board, who in May 2002 expressed alarm that students "know so little about their nation's history" and possess "so little capacity to reflect on its meaning" did a one-eighty eleven months later when rallying Congress for funds in history education:

Although it is customary for people of a certain age to complain about the inadequacies of the younger generation, such complaints ring hollow today.... What we have learned in these past few weeks is that this younger generation, as represented on the battlefields of Iraq, may well be our finest generation.¹⁵

The phrase "our finest generation" of course echoes the journalist Tom Brokaw's characterization of the men and women who fought in World War II as the "greatest generation." Those were the college students who in 1943 abandoned the safety of the quadrangle for the hazards of the beachhead. Yet only in our contemporary mirror do they look "great." Back then, grown-ups dismissed them as knuckleheads, even questioning their ability to fight. Writing in the *New York Times Magazine* in May 1942, a fretful Allan Nevins wondered whether a historically illiterate fighting force might be a national liability. "We cannot understand what we are fighting for unless we know how our principles developed." If "knowing our principles" means scoring well on objective tests, we might want to update that thesis.¹⁶

A sober look at a century of history testing provides no evidence for the "gradual disintegration of cultural memory" or a "*growing* historical ignorance." The only thing growing seems to be our

amnesia of past ignorance. If anything, test results across the last century point to a peculiar American neurosis: each generation's obsession with testing its young only to discover—and rediscover—their "shameful" ignorance. The consistency of results across time casts doubt on a presumed golden age of fact retention. Appeals to it are more the stuff of national lore and wistful nostalgia for a time that never was than a claim that can be anchored in the documentary record.¹⁷

Assessing the Assessors

The statistician Dale Whittington has shown that when results from the early part of the twentieth century are put side by side with those of recent tests, today's high school students do about as well as their parents, grandparents, and great-grandparents. That is remarkable when we compare today's near-universal enrollments with the elitist composition of the high school in the teens and early twenties. Young people's knowledge hovers with amazing consistency around the 40–50 percent mark—despite radical changes in the demographics of test takers across the century.¹⁸

The world has turned upside down in the last one hundred years, but students' ignorance of history has marched stolidly in place. Given changes in the knowledge historians deem most important, coupled with changes in who sits for the tests, why have scores remained flat?

Complex questions often require complex answers, but not here. Kids look dumb on history tests because the system conspires to make them look dumb. The system is rigged.

As practiced by the big testing companies, modern psychometrics guarantees that test results will conform to a symmetrical bell curve. Since the thirties, the main tool used to create these perfectly shaped bells has been the multiple-choice test, composed of many items, each with its own stem and set of alternatives. One alternative is the correct (or "keyed") answer; the others (in testers' argot, "distracters") are false. In the early days of large-scale testing, the unabashed goal of the multiple-choice item was to rank students, rather than to determine if they had attained a particular level of knowledge. A good item created "spread" among students by maximizing their differences. A bad item, conversely, created little spread; nearly everyone got it right (or wrong). The best way to ensure that most students would land under the curve's bell was to include a few questions that only the best students got right, a few questions that most got right, and a majority of questions that between 40 and 60 percent got right. In such examinations (known as "norm-referenced" tests because individual scores are compared against nationally representative samples, or "norms"), items are extensively field-tested to see if they behave properly. The testers' language is revealing: A good item is of medium difficulty and has a high "discrimination index"; students with higher scores will tend to get it right, and students with lower scores will tend to get it wrong. Items that deviate from this profile are dropped. In other words, only the questions that array students in a neatly shaped bell curve make it to the final version of the test.¹⁹

When large-scale testing was introduced into American classrooms in the 1930s, it ran counter to teachers' notions of what constituted average, below average, and exemplary performance. Most teachers believed that a failing score should be below 75 percent, and that an average score should be about 85 percent, a grade of B. Testing companies knew there would be a culture clash, so they prepared materials to allay teachers' concerns. In 1936 the Cooperative Test Service of the American Council on Education, ETS's forerunner, explained the new scoring system to teachers:

many teachers feel that each and every question should measure something, which all or at least a majority of well taught students should know or be able to do. When applied to tests of the type represented by the Cooperative series, these notions are serious misconceptions.... Ideally, the test should be adjusted in difficulty [so] that the least able students will score near zero, the average student will make about half the possible score, and the best students will just fall short of a perfect score.... The immediate purpose of these tests is to show, for

each individual, how he compares in understanding and ability to use what he has learned with other individuals.²⁰

The normal curve's legacy accompanies us today in the tests near and dear to the hearts of American high school students: the SAT (Scholastic Assessment Test). No matter how intelligent a given cohort of young people may be, no matter what miracles the standards movement may perform, no matter how much we close the "achievement gap" between students of different races and classes, it is impossible for the majority of students to score 1600. If that happened, the normal curve would not be normal, and Lake Wobegon, where "all of the children are above average," would not be fictional.²¹ It is impossible to have a basketball league in which every team wins most of its games. Similarly with the normal curve: there can be no winners without losers.



In spite of the odds against their scoring well, students at an SAT summer camp in Milton, Massachusetts, spend at least five hours a day studying and taking practice tests. *Courtesy National Center for Public Policy and Higher Education.*

If all students get an item correct, they do not necessarily know the material; the item's distracters may be lousy. Therefore, when we examine the names and events included among the distracters in the 2001 NAEP examination—for example, the mutiny of British forces under General Howe, the Wobblies, the Morrill Land Grant Act, and the relation between the coinage of silver and an economic downturn—we must remember that those facts do not appear because of their inherent worth or because they were taught in the high school curriculum or even because a blue-ribbon commission declared that every American high school student should know them. Rather, General Howe, the Wobblies, the Morrill Land Grant Act, and countless other bits of information appear on the test because they "work" mathematically; they snare their targets in sufficient numbers to boost the item's "discrimination index." It is modern psychometrics in the driver's seat, not sound historical judgment.²²

What would happen if the smart went down with the dumb, or, to put it more delicately, if students who knew the most history were stumped by items answered correctly by the less able? Again, the technology of testing makes sure that does not happen. Large-scale tests rely on a mathematical technique known as the biserial correlation in which each test item is linked mathematically to students' total scores and individual items that do not conform to the overall test pattern are eliminated from the final version.²³ Imagine an item about the *Crisis* magazine, which W. E. B. Du Bois edited, that is answered correctly at higher rates by black students than by whites,

while overall white students outscore blacks on the test by thirty points. The resulting correlation for the Du Bois item would be zero or negative, and its chances of survival would be slim—irrespective of whether historians thought the information was essential to test.²⁴ Technically, such examinations as the history NAEP are "standards-based" (hypothetically, every student should be able to "reach standard") rather than fit to a predetermined curve. But the practices of item analysis, discrimination, biserial or item-test correlations, and spread are so ingrained in psychometric culture that for all intents and purposes, results from all large-scale objective tests fit the traditional bell curve.²⁵ Indeed, this was confirmed by Steven Koffler, an administrator with the NAEP program that designed the 1987 history test, who reported that traditional item analysis and biserial correlations were used to create that supposedly standards-based test.²⁶

What does all of this mean, practically? First, in addition to handicapping students who possess different historical knowledge from those in the mainstream, it means that no national test can allow students to show themselves historically literate. If ETS statisticians determined during pilot testing that most students could identify George Washington, "The Star-Spangled Banner," Rosa Parks, the dropping of the bomb on Hiroshima, slavery as a main cause of the Civil War, the purpose of Auschwitz, Babe Ruth, Harriet Tubman, the civil rights movement, the "I Have a Dream" speech, all those items would be eliminated from the test, for such questions fail to discriminate among students. So, when the next national assessment rolls around in 2010, do not hold your breath for the headline announcing, "U.S. schoolchildren score well on the 100 most basic facts of American history." The architecture of modern psychometrics ensures that will never happen—no matter how good a job we do in the classroom.²⁷

Assessing the Future

Whether or not students have known history in the past begs our present concerns. Shouldn't we be worried when two-thirds of seventeen-year-olds cannot date the Civil War or half cannot identify the Soviet Union as our ally in World War II? No one concerned with young people's development should remain passive in the face of such results. Any thinking person would insist that such knowledge is critical to informed citizenship. In this sense, E. D. Hirsch, an educational critic and proponent of "cultural literacy," is right when he claims that without a framework for understanding—the ability to identify key figures, major events, and chronological sequences—the world becomes unintelligible and reading a newspaper well-nigh impossible. So why is that many young people emerge from high school lacking this core knowledge?²⁸

One narrative popular on both sides of the political aisle (as well as among historians who should know better) is that the social studies lobby and its agents warp young minds in ways reminiscent of a North Korean reeducation camp, wasting their time on mind-numbing "critical thinking" skills devoid of content. The problem with this armchair analysis is that, although we might find some support for it in the education school curriculum, the empirical data on what goes on in classrooms paint a different picture. Summarizing results from the 1987 national assessment, Diane Ravitch and Chester E. Finn concluded that in the typical social studies classroom, students

listen to the teacher explain the day's lesson, use the textbook, and take tests. Occasionally they watch a movie. Sometimes they memorize information or read stories about events and people. They seldom work with other students, use original documents, write term papers, or discuss the significance of what they are studying.²⁹

Similar conclusions emerged from a study of history and social studies instruction in Indiana in the early 1960s, as well as from John I. Goodlad's *A Place Called School*, the most extensive observational study of schooling in the twentieth century, involving twenty ethnographers in 1,350 classrooms observing 17,163 students. The high schools visited by Goodlad's team in 1977 all offered

courses in American history and government. Yet, while teachers in those courses overwhelmingly claimed that their goals fit snugly with "inquiry methods" and "hands-on learning," their tests told a different story, requiring little more than the recall of names and dates and memorized information. The topics of the history curriculum are of "great human interest," wrote Goodlad, but "something strange seems to have happened to them on their way to the classroom." History becomes removed from its "intrinsically human character, reduced to the dates and places readers will recall memorizing for tests."³⁰ Even at the height of the "new social studies" in the late 1960s and early 1970s, those who ventured into classrooms saw something different from the image conveyed by then-fashionable teachers' magazines. In the history classrooms Charles Silberman visited in the late 1960s, the great bulk of students' time was "devoted to detail, most of it trivial, much of it factually incorrect, and almost all of it unrelated to any concept, structure, cognitive strategy, or indeed anything other than the lesson plan." In a 1994 national survey of about fifteen hundred Americans conducted by Indiana University's Center for Survey Research (under the direction of Roy Rosenzweig and David Thelen), adults were asked to "pick one word or phrase to describe your experience with history classes in elementary or high school." "Boring" was the single most frequent description. Instruction lacked verve not because of projects, oral histories, simulations, or any of the other "progressive" ideas drubbed in print by E. D. Hirsch and others, but because of what the historian of pedagogy Larry Cuban has called "persistent instruction"—a single teacher standing in front of a group of 25–40 students, talking. A sixty-four-year-old Floridian described it this way: "The teacher would call out a certain date and then we would have to stand at attention and say what the date was. I hated it."³¹

Textbooks dominate instruction in today's high school history classes. These thousand-plus-page behemoths are written to satisfy innumerable interest groups, packing in so much detail that all but the most ardent become daunted.³² Many social studies teachers are forced to rely on the books because they lack adequate subject matter knowledge. Drawing on data from the National Center for Educational Statistics, Ravitch has shown that among those who teach history at the middle and secondary levels, only 18.5 percent possess a major (or minor!) in the discipline. Thus over 80 percent of today's history teachers did not study it in depth in college. While states have tripped over each other to beef up content standards for students, they have left untouched the minimalist requirements for those who teach history. The places that train teachers, colleges of education, are held hostage by accreditation agencies such as the National Council for Accreditation of Teacher Education (NCATE), which promote every cause, but not the stance that teachers should possess deep knowledge of their subject matter. ***Although federal efforts such as the Teaching American History program help individual teachers (and provide a boondoggle to many in the profession, including this author), at a policy level they are a colossal waste of money.*** One does not repair a rickety house by commissioning a paint job—one brings in a backhoe and starts digging up the foundation. Helping veteran teachers develop new subject matter knowledge for two weeks in the summer is a worthy cause, but not one likely to have any lasting impact. Change will occur only by intervening at the policy level to ensure that those who teach history know it themselves. Until that happens, we should expect no miracles.³³

What should we do while we are waiting for the revolution? First, we should admit that we cannot have our cake and eat it too. We cannot insist that every student know when World War II began and who our allies were while giving tests that ask about the battles of Saratoga and Oriskany. Today's standards documents, written to satisfy special interest groups and out-of-touch antiquarians, are a farce. When the majority of kids leaving school cannot date the Civil War and are confused about whether the Korean War predated or followed World War II, how far do we want to go in insisting that seventeen-year-olds know about the battle at Fort Wagner, Young-hill Kang's *East Goes West*, Carrie Chapman Catt, Ludwig von Mises, and *West Virginia State Board of Education v. Barnette* (Massachusetts State History Standards); John Hartranft (Pennsylvania State History

Standards); Henry Bessemer, Dwight Moody, Hiram Johnson, the Palmer raids, the 442nd Regiment Combat Team, Federalist Number 78, and *Adarand Constructors, Inc. v. Peña* (California History – Social Science Standards, grades 11 and 12); the policy of Bartolomé de Las Casas toward Indians in South America, Charles Bulfinch, Patience Wright, Charles Willson Peale, and the economic effects of the Townshend Acts (NAEP standards)? As William Cronon remarked after examining another pie-in-the-sky scheme, the ill-fated National History Standards, he would have been ecstatic if his University of Wisconsin graduate students had known half of this stuff. Or, in words that today's teenagers might use, let's get real.³⁴

None of the Above

The dilemmas we face today in assessing young people's knowledge differ little from those confronted by J. Carleton Bell and David F. McCollum in 1917. Few historians would argue that large-scale objective tests capture the range of meanings we attribute to the "historic sense." We use these tests and will do so in the future, not because they are historically sound or because they predict future engagement with historical study, but because they can be read by machines that produce easy-to-read graphs and bar charts. The tests comfort us with the illusion of systematicity—not to mention that scoring them costs a lot less than the alternatives.

Psychologists define a crazy person as someone who keeps doing the same thing but expects a different result. As long as textbooks dominate instruction, as long as the ETS dictates the history American children should know, as long as states continue to play a "mine-is-bigger-than-yours" standards game for students while hiding from view content-free teacher standards, as long as historians roll over and play dead in front of number-wielding psychometricians, we can have all the blue-ribbon commissions in the world but the results will be the same. Technology may have changed since 1917, but the capacity of the human mind to retain information has not. Students could master and retain the piles of information contained in 1917 or 1943 textbooks no better than they can retain what fills today's gargantuan tomes. Light rail excursions through mounds of factual information may be entertaining, but such dizzying tours leave few traces in memory. The mind demands pattern and form, and both are built up slowly and require repeated passes, with each pass going deeper and probing further.³⁵ If we want young people to know more history, we need to draw on a concept from medicine: triage. As the University of Tennessee's Wilfred McClay explains,

Memory is most powerful when it is purposeful and *selective*.... it requires that we possess stories and narratives that link facts in ways that are both meaningful and truthful, and provide a ... way of knowing what facts are worth attending to.... We remember those things that fit a template of meaning, and point to a larger whole. We fail to retain the details that, like wandering orphans, have no connection to anything of abiding concern.... The design of our courses and curricula must be an exercise in *triage*, in making hard choices about what gets thrown out of the story, so that the essentials can survive.... We need to be willing to identify those things that every American student needs to know and insist upon them ... while paring away vigorously at the rest.³⁶

Mechanical testing tempts us with the false promise of efficiency, a lure that whispers that there is an easier, less costly, more scientific way. But the truth is that the blackening of circles prepares us only to blacken more circles in the future. The sooner we realize this, the sooner we will be redeemed from our craziness.
