

This material may be protected by copyright law. (Title 17
U.S. Code)

DATE: 5, 28, 102
The History teacher.
Rivera D1 H61 Recent in Current Periodicals; Older in Stacks
LIB. HAS: v. 4(1970)-33(1999/00)-
Latest received: Feb 2002 35:2

D1. H61

|||||

Pending 20020528 Record 27 of 59
CAN YOU SUPPLY ? >YES >NO >COND >FUTUREDATE
ILL: 7440957 :Borrower: CUZ :ReqDate: 20020528 :NeedBefore: 20021201
:Status: PENDING 20020528 :RecDate: :RenewalReq:
:OCLC: :Source: Clio :DueDate: :NewDueDate:
:Lender: *CRU,CUS,CUT,UC4,UC4

CALLNO:
TITLE: The History Teacher
ARTICLE: Chu,, Jonathan "Preparing for the AP exam: the dangers of teaching
to the test"
VOL: 33 :NO: 4 :DATE: Aug 2000 :PAGES: 511-520
VERIFIED: OCLC: 1794874
PATRON: TELLEZ, KIP T./Crown Faculty Services
SHIP TO: University of California, Santa Cruz/McHenry Library/ILL/1156 High Street/Santa
Cruz CA USA 95064
BILL TO: same. fein# 94-1539563
SHIP VIA: US Mail :MAXCOST: |\$|20.00 IFM
:COPYRT COMPLIANCE: CCG
FAX: 831-459-3354, |vr|ARIEL PREFERRED|vn| Ariel: 128.114.163.19
BORROWING NOTES: DTD Yes, ktellez@cats.ucsc.edu, Account No: 22106001391902
LENDING CHARGES: :SHIPPED: :SHIP INSURANCE:
LENDING RESTRICTIONS:
LENDING NOTES:
RETURN TO:
RETURN VIA:

|||||

DATE: 5, 28, 102

Preparing for the AP Exam: The Dangers of Teaching to the Test

Jonathan M. Chu

University of Massachusetts-Boston

RECENTLY IN SAN ANTONIO, about 530 Educational Testing Service faculty consultants set about evaluating the essays of the over 176,000 students who took the Advanced Placement United States History examination. I had a special interest in two of the essays, having served on the Test Development Committee that drafted them and as one of those entrusted with the task of developing the standards or rubrics with which others would rank the essays for one of the questions. The results of this experience indicated to me a need to address how we prepare students to take standardized exams and to consider what we believe should be the objective of teaching history. In addition to eighty multiple choice questions, the AP exam asks each student to write three essays. One, the DBQ, asks the student to read a series of short documents and analyze them in response to a specific question. The other two essays (FRQs), each to be selected from a choice of two questions in two roughly topical categories, do not include documents, address a significant issue or subject in American history, are open-ended and are much like questions familiar to any student in a college level survey. In reviewing the answers to these questions, I found striking the large number of students who had written reasonable, well-crafted responses to two of them but had clearly misread the question. For example, in answering

FRQ #2 on the 1999 exam, one in six students, over fifteen times the usual number of students, responded inappropriately.¹ While these students clearly knew a good deal of American history, recognized some causal relationships, and wrote with reasonable clarity, their responses were wrong; they were plainly, simply, inappropriate answers.

What was most disturbing was that this and the other essay questions were not on obscure topics drawn from narrowly defined monographs crafted for arcane doctoral dissertations. Educational Testing Service faculty consultants, readers drawn from high school and university teachers, had expressed their general approval of what appeared to be student-friendly, straight-forward traditional questions. The DBQ used a series of documents whose authors ranged from Pennsylvanian Hector St. John Crevecoeur to Loyalist Mather Byles and asked students to evaluate “to what extent had the colonists developed a sense of the identity and unity as Americans by the eve of the Revolution?” The other essay which I will discuss, FRQ #2, asked: “How did two of the following—major political personalities, states’ rights or economic issues—contribute to the reemergence of a two party system in the period 1820-1840?”² Both were relatively mainstream topics—some might think of them as old-fashioned—staples of any college level U.S. survey.

My concern here is with the large numbers of students who responded to the DBQ with a discussion of the causes of the American Revolution and to the FRQ with a discussion of the early debates between Hamiltonian Federalists and Jeffersonian Republicans. On a 9 point scale, students who answered inappropriately received a 0, the same score they would have received if they had responded with the batting averages of the 1961 New York Yankees starting lineup. Many FRQ essays responded with extended descriptions of the differences in the personalities of Hamilton and Jefferson and in their disagreements over manufacturing and agriculture and over tariffs and monetary policy. A similar error occurred with respect to the DBQ.³ For example, with the overall task of evaluating the extent to which Americans had a sense of identity and unity on the eve of the Revolution, many students misused the list of contributions made by other towns and colonies for the relief of Boston, discussing the problem of taxation in the British Empire or describing the basic details of the Coercive Acts, rather than in using this and other documents provided to assess the degree to which the cause of Boston represented an emerging sense of identity or transcolonial unity. Other documents were similarly misused.

Whether one should partially compensate students for demonstrating some factual knowledge or analytical skill not entirely on point could be subject to debate;⁴ however, a more significant issue is why obviously able

students made these errors in such large numbers. At least three of these four issues, the causes of the American Revolution, the debate that gave rise to the first political parties or factions, or the rise of antebellum political parties are basic components of any United States history survey course—high school⁵ or university. The fourth—the existence of a sense of national unity or identity—is perhaps more demanding, a scholarly topic that is made complex by its utter simplicity; but it is arguably a question that lies at the heart of our history as a nation. Distinguishing among these events and arriving at a more appropriate response—and getting at least partial credit⁶—should not have been beyond the abilities of these students. Complex and sophisticated does not mean impossible, particularly for those who wish to be considered for college equivalency credit.

This specific kind of poor performance illustrates a shortcoming in the way in which many students were prepared for the examination and suggests the need for a more deep-rooted, reconsideration of the methods and goals of teaching history as we subject students to more standardized testing. Why did students discuss the early national period rather than the Age of Jackson or the causes of the American Revolution rather than the development of American nationality? The reason some students leapt to write on the wrong topic, unfortunately, may have been because of the way they were drilled in preparation for the examination. Teachers, not surprisingly or unwisely, tend to use as exercises the actual tests that have been given in previous years because these make the student familiar with the kinds of questions and issues raised. Moreover, with the issuance annually of the criteria for ranking, the standard or rubric, the teacher would have good guidelines for assessment which come directly from the people who construct the test. Not only do many teachers use past tests, but they also try to guess what the questions will be on the exam their students are preparing to take. The most susceptible to this process are the essays: the DBQ, for which ETS and the Test Development Committee provide a fifty year chronological restriction, and the free response questions. Guessing the DBQ topic allows teachers and students to concentrate their efforts in an area that will account for a significant proportion of the final ranking.⁷ Recently, one member of the Test Development Committee commented that when ETS announced the 1750-1800 designation for the 1999 DBQ some teachers indicated they focused their teaching on the period after 1750. Knowing the half century of the DBQ also allowed a narrowing of the range of anxiety for the free response questions⁸ because, by design, the FRQs must be drawn from fields not covered by the DBQ. Moreover, because students are afforded a choice of one of two questions in the FRQs, preparation in them becomes secondary.

The point we have been making about why so many students wrote inappropriate answers on the 1999 DBQ and FRQ #2R can be shown by looking further at the DBQ of the previous year. It had asked to what extent Jeffersonians could be characterized as strict, and Federalists as broad constitutional constructionists in the period 1801-1817. When ETS announced that the 1999 DBQ would be located in the period 1750-1800, it led to the not unlikely conclusion that it would be about the American Revolution or the Constitutional period. Leaping to the conclusion that the DBQ would be on either topic was low-risk. But if the student was encouraged to prepare rote-like, an essay on the causes of the Revolution or the ratification of the Constitution to the exclusion of any other possibilities, he or she was ill served. The class should certainly have discussed at some length the causes of the American Revolution. This would have been extremely useful in preparing an essay on the emergence of colonial unity or national identity, but the emphasis should have been not on perfecting a single essay but rather upon training the student to see that the same body of factual material could serve a number of different analytical paradigms or that the question might just be different. Students leaped to an inappropriate response because, by design or by implication, they had been trained to use a set of facts for a singular explanation, not to the goal of most university history teaching, of understanding how disparate historical facts woven together provide multiple explanations for events. Preparing a singular answer based upon this premise also explains the disastrous showing on the FRQ where students substituted the debates between Federalists and Democratic Republicans (the first party system) for a discussion of the emergence of the second in 1820-40. Students leaped to terms like personalities (Hamilton and Jefferson), economic issues (manufactures and agriculture), sectional issues (tariffs and economic issues again), but they did not have sufficient command of the facts to recognize that this question needed a new body of material that could be placed in an analytical construct with which they were familiar.

For all the best reasons, teachers may have encouraged students to use the 1998 DBQ as an exercise in crafting an essay. Because it provided some documents, because ETS distributes its criteria for evaluation after the 1998 reading and evaluation of the essays, students would have had an important model of a strong essay. Using a DBQ would have demonstrated an important lesson when undertaking an analysis. It would have given students historical documents and, with them, prompts that could lead them to developing an organized, analytical concept that in turn would have served as a frame for writing an effective essay. That so many students presented finished reasonably polished essays on Hamilton and

Jefferson in place of one on the Age of Jackson points again not to a lack of factual detail but to inadequate analytical training. Reversing the problem seen in the 1999 DBQ, students found themselves with a proper construct for an essay, but were unable to discern the need to substitute new facts to reorient it and bring it back on point.

The poor showing of students on the DBQ and the FRQ of the 1999 AP U.S. History examination demonstrates that many students were unable to use facts appropriately and reveals a more telling observation about the way in which we teach history generally and the AP course specifically. With regard to the FRQ, any reasonably diligent student who presented a coherent discussion of the first debates over Hamilton's reports on public finance or manufacturing and of Jefferson's agrarianism should have been able to say something about politics in the period 1820-1840 in a manner that should have received at least partial credit. Asked to address two of the following topics, major personalities, states' rights or economic issues, students should have been able to recall at a minimum two pieces of a large and, I would expect, familiar body of information to respond adequately in a number of different categories to the question asked—Andrew Jackson, Henry Clay, John C. Calhoun, Daniel Webster, Nicholas Biddle, the Compromise of 1820, the Tariff of 1828, the Bank of the United States. Students who could intelligently discuss the causes of the American Revolution should have been sufficiently skilled to recast their essays from a general description of causes to evidence of American identity or unity when given Crevecoeur's statement about "the American, this new man," or the section of the Continental Congress's "Declaration for the Causes of Taking up Arms." The addition of commentary by Loyalists Mather Byles and Peter Oliver which was supplied in the documents should have prompted responses about American identity and a sense of unity, even responses which distinguished the excellent from the satisfactory essay.⁹

While the failures of students to adjust their essays to these new demands raises questions about the way AP courses are sometimes taught, it also raises concerns about the possible misuse of AP examinations to measure teaching effectiveness or the quality of schools, a misuse which can cause excessive teacher concern about teaching to the test. These issues arise whenever standardized tests are misused for these purposes or when they seek to impose on students what authorities believe to be the right facts of the nation's history because they fulfill certain socializing purposes.¹⁰ Correctly taught and correctly used, however, tests like AP provide a perfect example of the effectiveness for standardized testing where the goal is specific, limited and validated. Over time with the help of the collective wisdom of many, with enormous

resources devoted to testing the statistical validity of each year's version, AP provides as close to a consensus as possible as to what a typical university-level U.S. history course should be. That is after all, the objective of the Educational Testing Service—to assure American colleges and universities that high school students from Portland, Maine and Portland, Oregon or from Frankfort, Kentucky and Frankfurt, Germany with the same score will demonstrate comparable levels of achievement in American History. And AP works: it purports to test college level achievement in American history, its results have a reputation for a high degree of statistical validity, and students who receive a three or higher are deemed to be qualified for college credit. Many colleges agree by granting those credits, though some insist upon a score of four. By extension then, a record of achievement—that is significant numbers of high school students receiving threes or higher—denotes university-level achievement in the secondary school, an easy and logical intimation of excellence in the classroom and excellence in the teacher that in turn speaks to the quality of the school.¹¹ When presented in standardized forms, they also give rankings that help us compare students' achievements across time and place.¹²

We have seen, however, that poor teaching practices can prepare students to do poorly on AP essay questions. It is also possible that excessive concern about having students do well on AP exams for teachers to focus unduly upon formulaic responses reflecting "objective" fact accumulation. Insisting that students know more historical facts is easy. Facts are discrete and short; we can verify them; they are easy to grade; who, when and where attributes are ascertainable. Given sufficient warning, teachers can then take students and drill into their short-term memory sufficient number of facts to facilitate achievement on the standardized test. But this reflects an objective inconsistent with what history, I think, offers liberal education.

With respect to AP U.S. History, teachers consistently bemoan the lack of time for preparation so that they can cover the material. They repeatedly stress the need for structural devices either in the exam or the testing process to give students more time to master facts. Teachers hold extra review sessions; they assign summer reading, ETS is asked to delay the date of the examination, it is asked to release recent multiple choice questions, and the Test Development Committee is urged to provide chronological cues or bullets for the DBQ or free-response essays or to structure the questions so a response using the appropriate facts is all but inevitable. While I see this as somewhat wrongheaded, this does not mean that I consider knowing facts is unimportant; indeed, I wonder why the conversation about content occurs or takes on such heat. Does anyone

ask the professor in the elementary physics class if knowing multiplication or division is important; or the English professor, if words matter? Knowing historical facts is like dribbling in basketball. Bob Cousy, Michael Jordan and Wilt Chamberlain all had to know how to dribble; but not all of them used it for the same purposes: Cousy to control the tempo of play, Jordan, to score; Chamberlain, when he had no other choice. I tell my students that the more facts they know the easier it is to see the relationships; I tell them that we all do not remember the same facts or the same things about the facts that we know in common, but that the sounder the historical explanation, the more seemingly disparate facts will support it. The objective is a more inclusive explanation that provides a better perspective on the human condition as it changes over time; achieving this frequently requires taking different routes. I emphasize this because, as we saw in these 1999 essay results, the problem does not lie in the lack of the knowledge of facts but in a basic misunderstanding of their function in the study of history.

We need to teach our students that the telling of the story of the past is, by its nature, a process of giving meaning and significance to different facts. While the teller of the story has a responsibility to be historically accurate, to root his or her discussion in facts that are ascertainable, he or she not only controls the narrative but selects and favors some at the expense of others. If we have not taught them this, then we have denied students both an understanding of the processes of change over time and the critical thinking and analytical capacities that are unique to the discipline. If we have not taught them to look instinctively for multiple explanations of why things change over time or to question the use of facts in the construction of historical narratives, have we not deprived them of capacities for criticism and an ability to distinguish cant from wisdom?

What does this mean in practical terms—in preparing students to take the AP exam or in teaching them history? Fifty percent of the exam rests upon knowing three major topics—the material covered by the document based and free response questions, and in the case of the former, much of the factual knowledge one needs is provided with included documents and in case of the latter, the student has choices. Preparing for these essays assumes the development of analytical skills—of learning to take evidence in the form of documents and facts and weaving them into a narrative that can explain why events happened or policies came to be. We can teach this more effectively and efficiently if we focus upon the small, but important story and place it properly in the larger context of circumstance and event. Historians of the American Revolution have focused intently on smaller stories that incorporate the many factors that led to colonial discontent. Gross's Concord, Morgan's Stamp Act Crisis,

Labaree's Tea Party, Maier's crowds all provide small foci with rich detail of how colonial Britons came to be Americans. As individual narratives rather than a listing of events, they provide explanations and models that can see the universe in the small, but compelling story. Could not the focus upon the small story—while demanding that it be placed in the larger context—provide that window into the larger universe? Could not an extended discussion of Revere's illustration of the Boston Massacre sufficiently placed in context—connected to the larger events described in the above books—also provide a window into all the issues raised in these books? Students who either studied the Stamp Act, the Tea Party or Revere's engraving in close detail would also have the context for placing the ideas of the Declaration of Independence and the Constitutional debates. In learning one detail well, would they not also have an intellectual place and thus an easier time for knowing and contextualizing others?

Teachers are right to give students past AP U.S. History exams but not as models of content, but rather as prototypes of responses. This is not a small distinction. What should be central to the preparation of students is the understanding that some facts of history are rendered more important, more significant of commentary, not by intrinsic merit but by context and connection, by the action of the observer testing their applicability to a larger theme or context. Memorizing lists of more facts cannot substitute for the careful attention to the interrelationship of details, a deep knowledge of their significance, and the understanding of the multiple ways they can be rearranged in different explanatory narratives. I would prefer to have a student know that his or her \$20 bill provided evidence for the answer for the questions on political parties in the Age of Jackson as well as ones on Civil War finance and Populism rather than one who could recite the ad valorem rates of the Tariff of Abominations, knew where John C. Calhoun is buried or who proposed the theory of bimetallism.¹³

Students need to understand that the most effective use of their factual knowledge comes through seeing the contingent uses to which that knowledge can be applied as they study the changes in America over time, to see the multiple relationships that exist in human history. They need to know facts so well that they reflexively look for multiple layers of analysis and instinctively test the use of the evidence from which they were derived. If we have taught them these lessons, we will have given them the most important intellectual gifts that history brings to critical thinking and learning. Just as Cousy, Jordan and Chamberlain understood that dribbling was important only in the context of winning basketball games, students must come to see that remembering facts alone is only one part of learning history.

Table 1
Distribution 1999
DBQ and FRQ 2¹⁴

<u>Ranks</u>	<u>DBQ</u>	<u>FRQ</u>
9	0.5	0.8
8	2.0	2.8
7	4.4	5.3
6	8.1	7.8
5	16.8	12.4
4	23.9	13.4
3	24.8	14.0
2	15.1	14.6
1	3.7	13.6
<u>0</u>	<u>0.7</u>	<u>15.6</u>
Totals	100.0	100.0

Notes

1. 15.6% of the students received a 0—an ETS ranking for an inappropriate response. The usual percentage is generally less than 1%.
2. See Table 1.
3. A student writing the DBQ has obvious advantages because he or she is able to refer to the attached documents. Since the documents all related to issues associated with the Revolution, exam graders probably were less likely to see the stark degree of inappropriateness of a student essay than was true of the FRQ. The nonresponsive rates for the DBQ was considerably lower than for the FRQ, 0.7% (including both O and/or NR) as opposed to 15.6%.
4. Obviously a case could be made for either position. One might argue that the student should receive partial credit for demonstrating some knowledge, skill or ability. Alternatively, most faculty consultants argued that students need to learn to respond directly and pointedly to the questions asked of them. I suspect the Educational Testing Service response would also argue that this would complicate exceedingly the statistical rankings from one year to the next.
5. Except, of course, in Massachusetts where new curriculum frameworks do not provide for teaching U.S. history before 1865 in high school. See Massachusetts Department of Education, *History and Social Science Curriculum Frameworks* (Malden, MA, 1997), 53-54.
6. I emphasize the partial credit and not college equivalency here. In 1998 the mean scores on the free response essays ranged from 3.28 to 3.84 on a 9 point scale;

overall, generally well over 50% and as much as 64% of the essays fall below 4. Statistics for 1998 were provided by the Educational Testing Service. Essays are given a ranking on a scale of 0-9. The scores from the essays are then combined with one for the multiple choice section and a score based upon a 5 point scale given for the overall test, with 3's ("qualified") and above, depending upon the college or university granting college equivalent credit. See The College Board, *Advanced Placement Program Course Description: History, United States History, European History, May 2000, May 2001* (Princeton, NJ, 1999), 76, or "The Acorn Book."

7. The essay and multiple choice sections each account for half of the total score. Within the essay section, the DBQ accounts for 45% and the two remaining FRQ, 55%.

8. There is some continued confusion on this particular point. Despite being told the contrary, some teachers remain convinced that the Test Development Committee and/or the Acorn Book, the official guide to the Test produced by ETS, advises that the two sections corresponds to a division between pre- and post-Civil War history. The sole guide to the development of the FRQ is that the essay section collectively represents a broad cross-section of chronological period and subject

9. "Lest this declaration should disquiet the minds of our friends and fellow-subjects in any part of the Empire, we assure them that we mean not to dissolve that union which has so long and so happily subsisted between us (*Declaration for the Causes of Taking Up Arms*)." "They call me a brainless Tory; but tell me my young friend, which is better, to be ruled by one tyrant three thousand miles away, or by three thousand tyrants one mile away (Mather Byles)."

10. I have purposely not engaged the problems created by the desire of senior administrators and policyholders to evaluate the quality of teachers through the standardized testing of their students.

11. For example, see Jay Matthews, *Class Struggle, What's Wrong (and Right) with America's Best Public Schools* (New York, 1998).

12. A crucial part of the test development process is the use of questions—equators—that help to make tests in different years statistically comparable. There are also ways of correlating the rankings of essays with correct answers on the multiple choice sections that statistically help to assess the overall levels of comparability of tests from one year to another.

13. Jackson, Specie Circular, greenbacks and National Bank (from the fact of paper currency).

14. Statistics have been provided by the Educational Testing Service.

AP U.S. History: Beneficial or Problematic?

Maxine N. Lurie

Seton Hall University

FOR SEVEN OF THE LAST EIGHT YEARS I have graded AP United States History exams at Trinity University in San Antonio, Texas, serving for the last three years as a "table leader" supervising other graders. This is a terribly important enterprise for all who teach United States history, whether on the high school or college level. But I have become more disturbed each year by what I see as problems, and suspect that to ignore them in the long run will imperil the program itself. The problems include too many exams (some taken inappropriately), which causes difficulties in maintaining standards. AP courses are meant to provide college level work, including not just the "facts" but also learning how to think historically, use primary source documents, and write coherently—instead too often they become lessons how on to take THE TEST. There are also problems in attracting graders, which reflect attitudes towards what does and does not count as "professional activities" for historians, particularly those teaching at the college level. What started out as a means of selecting the best and brightest of high school students, putting them through a rigorous history course, and rewarding them by permitting them to take an exam which could place them out of college level courses, is becoming a test of minimum standards that reputable colleges may soon ignore.