Preparing Teachers to Teach Historical Thinking?
The Interplay Between Professional Development Programs and School-Systems’ Cultures

Bruce VanSledright, Liliana Maggioni, & Kim Reddy

University of Maryland, College Park

April, 2011

Preparing Teachers to Teach Historical Thinking?

The Interplay Between Professional Development Programs and School-Systems’ Cultures

What is the American public getting for its investment in the Teaching American History (TAH) professional-development grant program? The answers to this multi-million dollar question vary, depending on whom you talk to, although we suspect that a common denominator across these responses would be that we simply do not know for sure. The reason of such uncertainty is linked to at least two matters that concern us here: (a) the absence to date of a comprehensive, widely disseminated federal report on the what the program has yielded in transforming American history teaching since its inception in 2001-2002, and (b) the relative dearth of published or disseminated reports from program evaluators about what their data suggest regarding the program’s impact on history teachers and teaching practice.

This paper intends to begin addressing, in part, the second concern. Regarding the first concern, the absence of a comprehensive, 10-year federal report is in good measure understandable. Because of the vast array of different program evaluation designs used by external program evaluators, the U.S. Department of Education has experienced considerable difficulty conducting a synthesis of results that could be reported out in meaningful ways, although they continue to work at the task (e.g., Humphries, et al., 2005). This leaves it up to researchers and program evaluators to attempt to fill the void.

Lurking behind these difficulties lies what we believe is a more fundamental question about the nature of the transformation hoped for as a result of teachers’ participation in professional development experiences made possible by TAH grant. Is it a rise in students’ achievement, and, if so, what is it that students are supposed to achieve? With respect to this latter
question, it is important to note that, in the United States, recurrent disappointment with achievement test results (e.g., Lee & Weiss, 2007; Rothstein, R., 2004; Wineburg, 1996), showing students’ inability to recall traditional milestones of American history, has often been met by a renewed commitment to foster students’ memorization of increasingly vast curricula. That effort attempts to conciliate the goal of introducing new generations of Americans to a common narrative of the nation’s past while paying at least some attention to the diverse voices characterizing American society (VanSledright, 2008).

Besides issues of curriculum coverage and memorization, the instance that learning history implies developing competence in the ways of thinking typical of this domain (Lee, 2004; Seixas, 2000) clearly (and controversially) entered the public debate, too, culminating in the revision of the National Standards for History in 1996, and the consequent adoption of this framework by several states developing their own state standards. The modification of the curricula that often ensued made professional development for teachers especially urgent. Yet, curricular change takes flesh within specific school systems, where decisions are made about what topics to cover, and what aspects of the process that generated those topics in the first place to uncover. Likewise, teacher professional development does not happen in a vacuum, but in the context of particular school systems’ cultures, each with its own way of envisioning the role of history in the school curriculum, the goals that it should pursue, and a way of framing and navigating the tensions that often arise between covering and uncovering history (sometimes framed as a tension between content and process).

This study examines and compares the results of three nearly identical professional development programs implemented with the support of Teaching American History grants. Specifically, it focuses on results from these programs’ efforts to reshape how the history teachers
they worked with think about and teach history to their students. In other words, we are especially interested in documenting what ideas about knowing history and learning history were addressed in these programs, how teachers reacted to them, and how they translated them into classroom practice.

The influence that teachers’ beliefs about knowing and learning have on instruction has been documented by several studies (e.g., Brickhouse, 1990; Hashweh, 1996; Zohar, Degani, & Vaaknin, 2001). At the same time, the influence that schooling in general and specific pedagogical approaches in particular may have on students’ epistemic development has also been documented in the literature (e.g., Elby, 2001; Elby & Hammer, 2010; Perry, 1970; VanSledright, 2002; Zohar & Nemet, 2002). Thus, in our study, we assume that the TAH teachers approach the professional development experience with heavily ingrained ideas about what history is, how it can be taught, and what they think their students are capable of learning of and about the subject. These ideas are shaped by their own long apprenticeships of observation in history classes both as once K-12 students themselves in history courses and as college students. Additional experiences as classroom history teachers in specific school systems continue to influence their ideas. The school history curriculum is a yet another influence, since it can further shape and reinforce ideas about what history is and how it needs to be taught (e.g., Paxton, 1999; VanSledright, 2008; Zimmerman, 2002).

Common school American history curricula, and the sort the teachers described in this study are asked to teach to everyday, center on the storyline of nation development and teachers report needing to tell that story using the standard textbook as the primary arbiter of the narrative register, cover its details, and instill the narrative in their students. In the TAH programs we have been studying and the three of focus here, partners (historians and master teachers) took a different
approach (or instructional theory), one more thoroughly rooted in disciplinary approaches, investigative in orientation, tailored around working with sources, reading the past, and writing interpretations, and more closely linked to the history education research literature (e.g., Donovan & Bransford, 2005; Seixas, 2000; VanSledright & Limon, 2006; Voss, 1998; Wineburg, 1996). As such, their view about what knowing history involves and what pedagogical practices may best serve student learning lack a direct alignment with school history orientations and instructional theories the teachers work from in their home schools. This conflict creates a serious intellectual and pedagogical hurdle for participating teachers as they encounter approaches to teaching history dissimilar to what they are being asked to do in their school. Cognitive dissonance can be a result and it can affect—positively or negatively—the influence of the program on their propensity to change.

This sort of cognitive dissonance involving education reform efforts in history is certainly not new. During the 1960s and 1970s, reformers on the Amherst History Project (Brown, 1996, 1999) encountered the same difficulties as they worked to reshape the teaching of history to make it cohere more closely with practices of coming to understand the past characteristic of the discipline itself. Those encounters left a legacy that some in the TAH program network have refashioned for the twenty-first century. However, the struggles that have ensued, such as those we describe here, bear uncanny resemblances to those of a half century ago. As Tyack and Cuban (1997) once wryly observed, no matter the results, we cannot resist tinkering toward some kind of educational utopia. As such, we attempt to update one story of this manner of tinkering for the new century, hoping to shine perhaps a slightly brighter light on the factors and difficulties involved, one that may assist other tinkerers.
Research Method

Contexts

The research contexts involved three large school systems in the mid-Atlantic region, all in the same state. Each was engaged in a federally-funded Teaching American History (TAH) grant program. The programs were designed by a group of historians at a local university and were constructed around the same professional-development model (bi-monthly sessions with teachers across the academic year followed by a week-long summer institute over a 3-year period of grant funding). The goal of the professional-development programs was to help history teachers rethink their typical history teaching practices, to stress the importance of investigating the past and teaching historical understanding through historical thinking as described in the research and reform literature (e.g., Donovan & Bransford, 2005; Levesque, 2008; Levstik & Barton, 1997; Stearns, et al., 2000; VanSledright, 2002; Wineburg, 2001).

We served as external evaluators on these programs, collecting a variety of data so as to inform project directors and professional developers of the successes and limits of their efforts. We provided data-based evaluation reports yearly for each of the projects. Observations were conducted of participating teachers and follow-up interviews with each were also undertaken. A multi-measure, paper-pencil assessment (called the Historical Knowledge Teacher Assessment, or HKTA) was constructed and administered in a time-series design in the context of each program, as an attempt to gauge change in participants’ epistemic beliefs, interests in history, and historical and pedagogical knowledge and thinking. We also attended and collected fieldnotes at all professional development sessions to document the nature of the interventions. These different types of data collection combined to help us understand the relations between the interventions.
and the potential for change in teachers’ historical ideas, beliefs, and thoughts about practices and their influence on what they actually did in their classrooms.

Participants

There were a total of 45 teacher participants that we followed. Although there were more teachers who enrolled initially in each of the programs, there was on average approximately 20% attrition across the three. We focus primarily on these 45 teachers because our data set on them is the most complete, providing opportunities to best gauge change over time. Seventeen teachers (slightly less than half were female) came from school system A (small urban core, large suburban population, some rural). Ten teachers (40% male) came from school system B (large suburban core wrapping around large post-industrial city, some rural population). Eighteen teachers (66% female) came from school system C (mostly a large exurban, bedroom community with a planned community at its center). In each group, teachers came from upper elementary (n=14), middle school (n=12), and high school (n=19) classrooms. Table 1 provides a breakdown of some of the other relevant demographic characteristics of each school systems’ teacher participants.

Table 1. Participant Demographics by School System

<table>
<thead>
<tr>
<th>School System</th>
<th>Number of Participating Teachers</th>
<th>Years Taught (mean)</th>
<th>Total US History Courses Taken by Participants (mean)</th>
<th>US History Courses Taken Last 5 Years (mean)</th>
<th>Average Number of History Workshops Participants Attended (last 3 yrs)</th>
<th>History Research Methods Course (percentage who had taken one)</th>
<th>Percentage of Teachers with Masters Degree in History</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>17</td>
<td>6.4</td>
<td>5.4</td>
<td>2.8</td>
<td>1.9</td>
<td>47%</td>
<td>18% (n = 3)</td>
</tr>
<tr>
<td>B</td>
<td>10</td>
<td>9.5</td>
<td>6.6</td>
<td>2.2</td>
<td>1.5</td>
<td>50%</td>
<td>0% (n = 0)</td>
</tr>
<tr>
<td>C</td>
<td>18</td>
<td>8.6</td>
<td>3.9</td>
<td>0.9</td>
<td>3.6</td>
<td>56%</td>
<td>11% (n = 2)</td>
</tr>
</tbody>
</table>
Years taught is an indication of the degree of relative teaching experience among participants. For example, teachers from A are typically younger and more inexperienced than both B and C. Total number of history courses taken, percentage who had taken a history research methods course, and the number of teachers with masters degrees in history suggest relative amounts of historical knowledge and background ideas participants had to work with entering the programs. Workshops attended indicate relative differentials in how much a school system engaged teachers in shaping their history teaching practices prior to the program introduction.

Measures

Three specific types of data inform our understanding for this paper. The first derives from and is embedded in the multi-measure, paper-pencil HKTA. Referred to as the BHQ (Beliefs about History Questionnaire), this measure samples participants’ level of agreement-disagreement on 22 statements that explore how historical investigators come to know the about past and how it can be taught to learners (see App. A). It yields an understanding of participants’ epistemic beliefs about history and ways of teaching it. It samples the degree to which participants identify with what we refer to as Copier, Subjectivist, and/or Criterialist beliefs (explained in more detail momentarily), and in the time-series design shows change (or lack thereof) over time (see Maggioni, et al, 2004; Maggioni, et al., 2009; Reddy & VanSledright, 2010).

Observational data from participants’ classrooms help shed light on (a) how practices reflected personal beliefs, (b) connections to how history was thought about and typically taught in the culture of schooling in the school systems (e.g., curricular requirements, standardized assessment, common teaching practices), and (c) changes in practices associated with ideas encountered via the TAH program interventions. We asked participants to let us observe lessons that they believed best represented the influence of the TAH program interventions. We then
engaged them in short interviews following these lessons. We queried them about their goals, their sense of the lesson’s efficacy (reasons for successes or lack thereof), what specific features of the TAH program they were intending to put in play, and any other ideas they wished to share with us.

Data Analysis

*BHQ Measure.* We analyzed the written responses to the BHQ adapting a method used in a previous studies to analyze data obtained from college students responding in writing to the BHQ (Maggioni, VanSledright, & Reddy, 2009; Reddy & VanSledright, 2010). In those studies, the researchers found that results obtained by using this method of analysis were compatible with results derived from qualitative analysis of students’ justifications of answers provided to the written BHQ. These justifications were offered in writing or during interviews.

Since items in the BHQ were written to reflect copier (items 5, 9, 16, 19, and 20), subjectivist (items 2, 4, 6, 8, 10, 12, 14, 17, and 22), and criterialist (items 1, 3, 7, 11, 13, 15, 18, and 21) beliefs (see App. A), a consistent epistemic position should produce agreement with items mirroring that specific stance and disagreement with items indicative of the other two stances. The first step in the analysis involved attributing values to the 6 response levels of the Likert scale used on the BHQ. Although we did not intend to assume continuity of the scale, relative numbers seemed particularly apt for the purposes of this analysis, because their sign could represent the position toward the statement (agreement or disagreement) and their value could represent the general strength of the decision (strongly or somewhat). In this way, we sought to maintain the order captured by the Likert scale, while creating a useful tool for the analyses described in what follows.
First, responses were scored using the following equivalencies: strongly agree = +3; agree = +2; somewhat agree = +1; somewhat disagree = −1; disagree = −2; strongly disagree = −3. Then, overall positions were assessed toward a certain stance by calculating weighted average scores; we did so by summing the scores obtained on the items reflecting that particular stance and dividing the result by the number of items mirroring that particular stance. For example, a respondent who agreed with item 5 (+2) and 9 (+2), somewhat disagreed with item 16 (−1) and disagreed with item 19 (−2), and strongly agreed with item 20 (+3) (all items reflecting a copier stance) would receive a score of 
\[
\frac{(+2)+(+2)+(-1)+(-2)+(+3)}{5 \text{ items}} = +1 \text{ (weighted score)}.
\]
A score with a positive sign (+) was interpreted as an indication that there was a degree of agreement with the copier stance and it tended to be stronger than the degree of disagreement. The value of the score (+1, in the example) is an indication that such agreement was overall moderate (equivalent of “somewhat agreement”). As such, these scores were used as provisional suggestions of the compatibility (or not) of a respondent’s beliefs with one of the theoretically derived epistemic stances (copier, subjectivist, criterialist).

Yet, this score does not indicate the consistency of the respondent’s stance. In fact, another respondent might obtain the same score (+1) by somewhat agreeing with all the items reflecting the copier stance and thus suggesting a more consistent copier stance. Because the TAH interventions (by structure and goal frameworks) were theoretically designed to enhance the criterialist beliefs of participants, we chose to assess epistemic consistency in relation to the criterialist stance. Beliefs characterizing the criterialist stance are also preferable because they are more reflective of the nature of historical knowledge and generally valued by experts (Maggioni et al., 2004). This, in part, explains the embrace of them by TAH project directors and professional developers given their intervention goals. A second reason was similarly important: In past
studies, respondents tended to agree with items mirroring the criterialist stance, thus suggesting a relative preference for it. Yet, we asked, how consistent was this preference? Developing a consistency score was an effort to gauge that preference.

Perfect consistency with the criterialist stance would be indicated by responses stating agreement (+1, +2, or +3 scores) on all criterialist items and disagreement (−1, −2, or −3 scores) on all copier and subjectivist items. To express the degree of consistency, we created a ratio (expressed in percentage) between the number of such responses and the total number of responses (22, if a respondent responded to all the items of the questionnaire). We called this ratio a consistency score. For example, a respondent agreeing with all the copier items (total = 5), disagreeing with all the subjectivist items (total = 9) and agreeing with all the criterialist items (total = 8) would obtain a consistency score of 55%, i.e., [(8 + 9 − 5)=12/22].

By school system, we calculated weighted mean scores for each stance and mean consistency scores. The time-series design of the BHQ measure allows us to track changes (deltas) in mean stance and consistency scores over time as the interventions proceeded. For all 45 participants, we calculated and tested correlations between weighted scores on three epistemic stances and consistency scores at the pretest and each successive administration to see, for example, if we could ascertain any patterns in the full data set. We used one-way ANOVAs to test whether teachers’ epistemic stances significantly differed across teacher grade levels, participation in history methods courses, and possession of masters degrees. We also calculated and tested correlations between weighted scores on the three stances and consistency scores, teaching experience, and history courses taken.

Although similar in school culture, beliefs, and structures, each school system is, however, somewhat unique in terms of how it pursues the U.S. history curriculum; how often, in what ways,
Preparing Teachers

and to what end it tests its students in history; the degree to which, for example, it stresses history programs such as Advanced Placement and International Baccalaureate; and the ways it pushes technical fixes (i.e., prescribes the history curriculum, draws from prepackaged curriculum materials) to mobilize student achievement in history. They also vary in the degree of top-down pressure teachers report experiencing from supervisors. We have observed many of these differences via program documentation, as they interjected their way into the three TAH programs. They can influence teachers’ beliefs and practices, thus our efforts to mine the BHQ data set for possible traces of these influences.

Observations and Interviews. In the observational and interview data, we looked at patterns of evidence for the tension between the vision of history and teaching practice evinced by the TAH program, and what teachers were doing in their classrooms and saying about what they were doing and why. As noted, teachers were asked to choose lessons for us to observe that they believed demonstrated the best of what they had learned in the TAH program. As a result, we were able to watch them attempt to practice that learning, obtaining a sense of the degree to which they experienced smooth transitions, or cognitive dissonance and tension, or had reinterpreted the program’s vision into something else.

We conducted on average two observations (gathering detailed fieldnotes) of each participating teacher across the three program years. Occasionally some teachers, despite our efforts, did not make themselves available to be observed and/or did not follow through with our requests for observation timeframes. For school systems B and C, we included in our analysis only those teachers for whom we had at least one pair of observations/interviews. While the TAH programs in these two school systems have been completed, the program in school system A is still operating and in its third year. To date, we have a total of 16 observations/interviews, some
teachers with two, others with one. Observations there continue and we do not yet hold a full complement of paired observations/interview for all teachers, but draw from the ones we do possess. We collected nine pairs of observations/interviews in school system B (18 total) and 14 pairs in school system C (28 total).

In analyzing these data, we initially read fieldnotes and interview data and commenced a discussion of potential themes emerging from that data. It was here where we first began noticing specific tensions between the TAH program goals and the exigencies of school culture. We generated a list of themes and then shaped them into codes following data examples and the theories of knowledge and teaching history they appeared to suggest. This process resulted in eight coding categories. Three derived primarily from the observations themselves (two levels—partial, full—of teaching disciplinary-type historical thinking to students, and, if not teaching historical thinking as understanding in lesson observed, then what instead). Five codes emerged from interview data and included (a) understandings expressed of historical thinking, (b) perception of goals of the TAH program, expressions of (c) epistemic stances and (d) school curricular goals, and (e) compartmental separation of the TAH program from activity in history classrooms.

In a second pass through observational and interview data, we applied the coding categories directly, establishing frequency counts. In an iterative cycle, we used this step to further refine the coding categories (adding one more, an “other” category for remarks not fitting previous codes). We further discussed and compared our codings and then the first author synthesized the results into frequency counts by category. The process allowed us a third pass in which we identified salient ideas, issues, and practices that spoke to the relationship between TAH
programmatic goal frameworks and efforts to influence the teachers and how they were interpreted by participants and applied in history classrooms.

Results

Because the TAH interventions were designed to help teacher participants rethink what history was as well as rethinking their teaching practices, it is important to realize, as we have alluded, that they ran headlong into the time-tested tension of trying to change sometimes deeply embedded beliefs that were underpinned by school-system based structures and attendant sanctions. That is, project directors and professional developers sought to frame history as a thoroughly interpretive enterprise that required the knower to wrestle with the objects of the past (using heuristics and criteria for making interpretive judgments) and therefore to teach historical understanding in schools as a slow-paced investigative effort that engaged students (as knowers) in efforts to interpret the American past through it’s residual objects (multiple accounts, artifacts, relics). They were met by teachers who were facing fast-paced curricular coverage mandates and supporting tests that had perennially produced teaching-as-telling practices and history as one-right story told in the standard U.S. history textbook (effectively a copy of the past). In many ways, the results obtained from our data analysis echo this tension and demonstrate its consequences for participating teachers.

As we also have suggested, despite the presence of clear similarities among the school system’s cultures and visions about U.S. history curricula and teaching, there were some subtle differences. We begin by noting in more detail these similarities and contrasting them with differences.
Schooling Cultures and History Education

All three systems were required to align themselves with the state’s specifications about the scope and sequence of the U.S. history curriculum. Because this scope and sequence tends to be, as they say, a mile wide and therefore an inch deep, school systems effectively must engage in curricular specifications that mirror it. As a consequence, history teachers in all three systems faced a battle against the clock to cover everything. By the end of the year, the teachers were often breathless and anxious, worrying about whether they dealt with it all adequately. The school systems differ a bit in what colligatory topics they stress. However, the differences are relatively minor. Some school systems in the state, although not these three, have attempted to manage this relentless pace by prescribing pacing guides for teachers, requiring that they be on such-and-such historical topic on day 27, then that topic on day 28, another topic by day 29, and so on across the school year’s 180 days. These three systems were less inclined to take this path but had considered variations on it, making less-prescriptive suggestions to history teachers.

Because of its exchange value and media-hyped cachet, all the school systems stressed the importance, for example, of recruiting as many students as possible into AP U.S. History courses and trying to provide support to enhance student success. Despite being, of course, different in some ways than the state’s U.S. history specifications and guidelines, current AP U.S. History also works from a coverage model. Some might argue that it functions like the “regular” U.S. history curriculum on steroids, as a few teachers have remarked. Of the three, school system A has done more to ratchet up the pressure on earlier-grades “regular” history teachers (middle school, for example) to prepare students for entrance into the AP course in high school. In fact, they have made repeated, clear public announcements about this goal. We have observed a number of non-AP TAH teachers in this school system teaching the AP A.P.P.A.R.T.S. acronymal and technical
learning strategy to students as early as eighth grade, as a means of ostensibly preparing them to be successful in later AP history courses. The other school systems are somewhat more subtle in their approaches to AP U.S. History.

Another similarity across school systems is their embrace of the practice (stressed in the TAH program) of using primary sources to understand the past. But all three systems still equip students with one version or another of the standard U.S. history textbook. When interpretive questions arise in the classroom, the textbook often serves as the arbiter of a dispute. At the end of the day, what we typically see across the systems is the common schooling practice of teaching to the one correct answer. Primary sources are, more often than not, pressed into this correct-answer pursuit. Classroom history, therefore, becomes less about open investigation of historical circumstances driven by rich questions than an exercise in establishing the right facts and predetermined interpretation supported by the textbook as well as the other sources. The coverage model supports and helps sanction this approach. Testing practices also aid and abet it.

Although the state does not pursue high-stakes testing in U.S. history, the school systems have developed what they call “district-wide unit assessments.” What unites the three systems is the stress on getting correct answers on the tests (likely following the spirit of the state curriculum standards) and that the tests are largely made up of multiple-choice and similar items that can be scored as right or wrong. Essay writing is also a similarity on such tests. The rubrics used to score them (sometimes borrowed from the state), however, tend to stress that scorers pay as much or perhaps more attention to generic and technical writing strategies and capabilities (e.g. correct grammar and spelling, 5-paragraph and thesis-support-conclusion structure) when they score essays than to more complex history-specific thinking and writing performances that could be scored on such essays. The exception here is a common rubric category for “supporting a position with
evidence.” But what this means exactly is rarely made clear or parsed for what it might entail exactly in an interpretive history essay. Here the systems are more similar than they are different.

Where they are distinguishable most measurably is how they guide, prescribe, and support their U.S. history teachers down this path toward helping students score better on the assessments. As one school person once remarked, she had never seen a technical approach that held even faint promise to facilitate higher test scores of which a school system disapproved. All three employed technical “fixes” of one sort or another, and there are too many of them to elaborate here. Suffice it to say, that especially among the elementary and middle school history teachers, the teaching of reading comprehension and its panoply of strategies as techniques (e.g., finding the main idea, rereading when comprehension breaks down, summarizing the author’s ideas) had been pressed into a high art. The AP APPARTS acronymal strategy matrix, as noted, was central to teaching history in school system A. For reasons unclear to us, we continuously observed teachers in that school system appear somewhat more predisposed to endorse any technical devices (acronyms, mnemonic strategy structures) they thought might be helpful. Acronym strategy stew was common, more common in school system A, but not uncommon in B and C.

The use of digital techniques for conveying history as information to be memorized were also common across school systems. The Powerpoint presentation was ubiquitous. However, there was some variation in how it was used. Sometimes it was employed as the digital version of the old analog chalkboard lecture notes or acetate transparencies. On other occasions, history teachers embedded digitized primary source documents in Powerpoint slides for classroom examination that facilitated the key points of a lecture. Occasionally, we observed the use of the student “clickers” system (mostly in secondary classes in school system A). Teachers would review a

1 Somewhat unhappy with these tests and the sorts of memorization-as-learning they promote, school systems B and C are now experimenting with different types of unit assessments that arguably have the potential to be much better aligned with history as an investigation into the past rather than a correct-answer pursuit arbitrated by the textbook.
previous-day’s lesson by posting true-false and multiple choice questions on the screen and students would attempt to “click” the correct answer, after which the teacher would tally the results, sometimes to students’ chagrin.

History as a discipline and an investigative activity of the sort TAH project directors and professional developers were trying to promote tends to defy easy technical fixes. There are investigative heuristics and highly valued evaluation criteria involved in coming to understand the past (Lee, 2005; VanSledright, 2011). But they tend to be wielded with considerable context- and question-driven judgment by expert investigators (Wineburg, 1991). As such, the TAH program approach in which teachers heard much about history as investigation into the past, and practiced doing it themselves on research projects as part of the program, rubbed uneasily up against variations on teaching-as-telling, history-as-correct-answer, fast-paced content coverage, and the embrace of often generic teaching techniques to facilitate better student test scores. A domain prizing professional judgment rooted in criteria meets the accountability culture. Six teachers across the school systems commented on their efforts to compartmentalize the two because they simply were unsure how to make their different efforts and approaches compatible. For these six, the TAH program and what they did in it was for them personally. What happened in school history was for school and for administrators to whom they were accountable and for the kids who needed to learn the story of their nation.

In what other ways did the three programs’ teachers navigate the tension produced between two differing approaches to history education? First we explore teachers’ epistemic beliefs. Did the TAH programs do much to change those beliefs? If so, in what ways? Since history-as-investigation (or understanding-the-past-as-historical thinking) appears to operate best with certain epistemic beliefs in mind, and that it can differ from history-as-story-memorization, these types of
beliefs are of considerable importance if the goal is to help teachers come to understand and work from the former idea.

**Epistemic Beliefs**

Three types of epistemic beliefs about history concern us here. The first is what we categorize as copier beliefs. In short, at its most naive, the copier tends to believe that the past and history are one in the same. The past happened as it actually happened and history simply narrates it, or perhaps it would be better to say, chronicles it. The objects of and from the past exist to anchor the story. And there is really only one correct way to tell that story—as it actually happened, as told to us by those objects. The knower here functions largely as a passive receiver. The principal problem for the copier occurs when the objects of the past (sources, relics) do not simply line up neatly to produce one correct story. There are competing and contradictory testimonies about a single event, for example. The copier has no real strategy for dealing with this common historical problem, and so when facing it, tends to get cognitively stuck, caught on a mental impasse that stops understanding.

The second is what we term the subjectivist. The subjectivist tends to be a bit of a disgruntled copier, one who, sometimes begrudgingly, has learned that in history the knower plays an important interpretive role in understanding the past. Objects do not speak in their own words; knowers have to interpret those objects. But the subjectivist, especially the more naive s/he is, tends to go too far in privileging the role of the knower. This typically occurs because the subjectivist lacks a rigorous set of epistemic anchors (judgment criteria) to manage the knower-objects interplay. That is, the subjectivist comes to believe that history is whatever we knowers make it to be. It’s all opinions all the way down. There are few clear methods for determining a better history from one less so. Forced to choose, the subjectivist punts, or picks one she likes (or
some combination of histories cut and pasted together), and in the face of criticism, defends the choice by invoking the rule that in democracies, everyone’s entitled to her or his opinion. It’s history as anything goes. And again, understanding stops, because short of learning key judgment criteria and analytic heuristics, one cannot arbitrate among various histories defensibly.

The third is what we term the criterialist. The criterialist believes that deeper historical understanding is possible when we can obtain ways to coordinate the balance between the objects of understanding and the subject. Both play a role. The knower, or subject, must be constrained by the nature of the objects, the contexts in which they were generated in the first place, the questions that can be reasonably asked of them, and the limited range of claims they make possible. But criterialists also operate as knowers on what can be known. They work from carefully honed analytic strategies and heuristics. They prize power of judgment, but that judgment is constrained by the objects and by the community in which they converse about the past. The problem of coordination between knower and objects cannot be solved completely, but the criterialist possesses the tools to manage it so that understanding can go forward. Historian James Kloppenberg (1989) summed up the belief structure this way: “Beyond the noble dream of scientific objectivity and the nightmare of complete relativism lies the terrain of pragmatic truth, which provides us with hypotheses, provisional syntheses, imaginative but warranted interpretations, which then provide the basis for continuing inquiry and experimentation. Such historical [reading and] writing can provide knowledge that is useful even if it must be tentative.” (p. 1030)

The TAH programs in each school system worked off an expressed goal to prepare the teachers to think more like historians, to wrestle with the past and its multiple possible meanings in the way Kloppenberg (1989) suggests. The hope was that if the teachers could learn to think
this way, they could teach their students to do the same. Students’ understanding and therefore achievement would increase. That in short was the theory guiding the interventions. Without directly using the vocabulary of copiers, subjectivists, and criterialists, the TAH program directors and professional developers (historians, master teachers drawn from each school system working together) were, nonetheless, seeking to enhance criterialist epistemic beliefs and reduce copier and subjectivist beliefs. They recognized that copier and subjectivist beliefs tended to be unproductive in deepening historical understanding that they were asking of the teachers in both (a) the history research projects they were required to undertake in the programs and (b) in applications back to school classrooms.

We tracked these beliefs to see if we could detect movement. As we have described, the BHQ was the key tool for doing so. We used data from observations and interviews to triangulate results and explore how teachers’ beliefs surfaced in the way teachers talked about the relations among the program, their practice, and in their work with students in their classroom. Prior studies have indicated that, with the exception of experts, individuals rarely voice beliefs that neatly align with one of theoretical stances described above (see Maggioni, 2010; Maggioni, et al., 2004). Rather, epistemic inconsistency appears to be the norm. The consistency score we calculated for each teacher is a means of assessing the degree to which teachers are consistent in their beliefs with the criterialist position. The higher the consistency score (expressed as a percentage), the more teachers on average move toward criterialist epistemic beliefs and away from copier and subjectivist beliefs. Table 2 depicts mean consistency scores for each school system’s teachers over four administrations of the BHQ in school systems B and C, and three in A.

Consistency changes across programs suggest that, for teachers in school systems A and C, the pattern from first administration to last (or beginning of the program until the end in C and
until the end of year 2 in A) is that teachers move slightly toward becoming more solid.

Table 2. Consistency Scores by School System TAH Teachers Across BHQ Administrations

<table>
<thead>
<tr>
<th></th>
<th>School System A</th>
<th></th>
<th>School System B</th>
<th></th>
<th>School System C</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N=      Mean  SD</td>
<td>N=      Mean  SD</td>
<td>N=      Mean  SD</td>
<td></td>
<td>N=      Mean  SD</td>
<td></td>
</tr>
<tr>
<td>Consistency 1</td>
<td>17      .81  .08</td>
<td>10      .78* .10</td>
<td>18      .77  .11</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Consistency 2</td>
<td>17      .82  .10</td>
<td>10      .64* .00</td>
<td>18      .76  .07</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Consistency 3</td>
<td>17      .83  .12</td>
<td>10      .79  .11</td>
<td>18      .76  .08</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Consistency 4</td>
<td>–       –      –</td>
<td>10      .78  .14</td>
<td>18      .79  .09</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note: t-Tests of statistical significance on mean score changes from BHQ administration to administration showed a few instances of significance (p ≤ .01), which are indicated by asterisks.

criticalists. Teachers in B ended where they began, after a large drop in consistency at the end of year 1, a rebound, and then a very slight decline by the end of year 3. However, since the differences between consistency scores at the beginning and at the end point of the program were statistically insignificant, we cannot make claims about the generalizability of the changes with much confidence and we limit ourselves to observing general patterns.

Of potential relevant note, the TAH program in school system B was the oldest of the three (2005-2008), followed by C (2007-2010), and then A (2009-ongoing). It is possible to argue that professional developers, many of whom worked across programs over the years, became more astute at influencing teachers’ epistemic beliefs in the directions intended over the years of the programs. This could partially help explain the steadier rise in teachers’ consistency scores in A, and the final rebound in the last year among teachers in C. On the other hand, in other studies using the BHQ among college students (Reddy & VanSledright, 2010), we found that consistency often drops for a period once the intervention begins, as it challenges some epistemic assumptions and creates early uncertainty. This appears to be a similar pattern among teachers in C and to a lesser degree in B. However, teachers in school system A defy this pattern for reasons that remain unclear to us.
We tested the statistical significance of mean differences across school systems in consistency scores at the beginning and at the end of the program and in year-to-year deltas in consistency scores, but the results were all statistically non-significant. It may well be the case that individual differences among teachers in terms of the structure of the epistemic beliefs with which they entered the program, and in terms of their patterns of change over time, play a bigger role than teachers’ belonging to a specific school system in explaining the epistemic influence (or lack thereof) of the programs. This result may also be a by-product of the relative similarity across these school systems in terms of factors that may hinder or foster teachers’ movement towards criterialist beliefs.

In our previous analyses of BHQ consistency scores (Maggioni, et al., 2009; Reddy & VanSledright, 2010), we elected to set the bar at or above 90% (.90) consistency in order to say that someone had established a strong criterialist stance. Examining mean scores here, on average teachers in all three school systems fell short of that bar, suggesting that after three years in the TAH programs in the case of school systems B and C, and two years in A, teachers on average remained relatively inconsistent in their epistemic stances (although, across systems, there were a few individual teachers who did achieve strong criterialist scores and were quite consistent in them). In other words, they continued to vacillate between the three positions. As we will see, this was more so the case with variations on the relationship between subjectivism and criterialism. Overall, the TAH programs appeared to have difficulty moving teachers toward a more consistent and robust coordination of beliefs about the role of the knower vis-a-vis the objects of the past that characterize a criterialist position. The role of school system culture—one that at least subtly encourages copier beliefs—could again be playing a role here.
To examine movement and consistency more closely, we calculated each teacher’s score on each theoretical epistemic position—copier (Co), subjectivist (Sub), and criterialist (Cr). We then computed weighted mean scores on each position by school system, and again ran t-tests to compare mean score changes from year to year on the programs. Tables 3 through 5 are provided sequentially to display data for this comparison, noting statistical significance on mean score changes from BHQ administration to administration (<.01 or <.05) by asterisks. Negative signs indicate disagreement with the position. The range is –3 to –.01 for disagreement, 0 for ambivalence, and +.01 to +3 for agreement with the higher numbers (e.g., +3, –3) representing stronger agreement/disagreement.

Table 3. Teachers’ Epistemic Position Weighted Means by Year (N = 17)

<table>
<thead>
<tr>
<th>School System A</th>
<th>Copier (Co)</th>
<th>Subjectivist (Sub)</th>
<th>Criterialist (Cr)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>SD</td>
<td>Mean</td>
</tr>
<tr>
<td>BHQ 1 (Pre-program)</td>
<td>–1.1</td>
<td>.98</td>
<td>–.93*</td>
</tr>
<tr>
<td>BHQ 2 (End Year 1)</td>
<td>–1.1</td>
<td>.91</td>
<td>–.89</td>
</tr>
<tr>
<td>BHQ 3 (End Year 2)</td>
<td>–1.5</td>
<td>.78</td>
<td>–1.2*</td>
</tr>
<tr>
<td>BHQ 4 (End Year 3)</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
</tbody>
</table>

Table 4. Teachers’ Epistemic Position Weighted Means by Year (N = 10)

<table>
<thead>
<tr>
<th>School System B</th>
<th>Copier</th>
<th>Subjectivist</th>
<th>Criterialist</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>SD</td>
<td>Mean</td>
</tr>
<tr>
<td>BHQ 1 (Pre-program)</td>
<td>–1.1*</td>
<td>.72</td>
<td>–.93*</td>
</tr>
<tr>
<td>BHQ 2 (End Year 1)</td>
<td>–1.9*</td>
<td>.25</td>
<td>–2.1*</td>
</tr>
<tr>
<td>BHQ 3 (End Year 2)</td>
<td>–1.4</td>
<td>.68</td>
<td>–.68*</td>
</tr>
<tr>
<td>BHQ 4 (End Year 3)</td>
<td>–1.0</td>
<td>.87</td>
<td>–.71</td>
</tr>
</tbody>
</table>

Table 5. Teachers’ Epistemic Position Means by Year (N = 18)

<table>
<thead>
<tr>
<th>School System C</th>
<th>Copier</th>
<th>Subjectivist</th>
<th>Criterialist</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>SD</td>
<td>Mean</td>
</tr>
<tr>
<td>BHQ 1 (Pre-program)</td>
<td>–1.5</td>
<td>.95</td>
<td>–.38</td>
</tr>
<tr>
<td>BHQ 2 (End Year 1)</td>
<td>–.79</td>
<td>2.51</td>
<td>–.58</td>
</tr>
<tr>
<td>BHQ 3 (End Year 2)</td>
<td>–1.2*</td>
<td>.80</td>
<td>–.57*</td>
</tr>
<tr>
<td>BHQ 4 (End Year 3)</td>
<td>–1.7*</td>
<td>.71</td>
<td>–.73</td>
</tr>
</tbody>
</table>
As a whole, teachers in all three systems were less attracted to copier beliefs, although weighted averages indicated that, for the most part, they only somewhat disagreed with those items. There were exceptions: System B teachers at the end of the first year fully disagreed (–1.9) with the copier stance and system C teachers approached that full disagreement by the end of the program. With system B and C teachers, we again see some vacillation in the middle of the programs. Teachers in system B became less agreeable to copier belief items and then returned to their somewhat disagree position by the end. Teachers in system C demonstrated the reverse pattern by growing less disenchanted with copier beliefs in the first two years of their program and then becoming more disagreeable to them by the conclusion.

Subjectivist beliefs on average are more attractive to the teachers in all three systems. Although nearly all the weighted mean signs are negative, showing some disagreement, they are closer to the zero position of ambivalence than are copier beliefs. This seems understandable theoretically. Because history is a thoroughly interpretive discipline, the role of the knower is important since it is the knower who fashions the interpretation. Despite criteria (e.g. reliance on evidence preponderance to make claims) that constrain the knower, the knower (or subject) is still largely in control of how those criteria are utilized and as a result what that interpretation looks like.

What these BHQ data do not fully tell us is the degree to which the teachers are attracted to a relatively naive “anything-goes,” “free-speech-rights” version of history, or a more reasoned and limited version that anticipates the criterialist position. Our observation data in particular seem to suggest that it is more the former, given how often they used the word opinion in favorable ways when talking about interpretations. Teachers in general were not able to learn enough about how to coordinate their subjectivist leanings with history’s criterial judgment under the relatively short
duration of the programs (and to date our data set for system A teachers remains partial).

Moreover, as we have seen, much about the culture of schooling in the systems and history education in them work against the embrace of a fully coordinated, consistent set of beliefs that favor a criterialist stance. In fact as stressed, they tend to press teachers in the opposite direction, toward copier beliefs that seem to underpin history education visions there (e.g., the correct, tested details and interpretation are in the textbook, history and the past are isomorphic). At the same time, whenever teachers did open up some space for students’ voices through the use of discussions and debates, few suggestions were made about rules for argument that would guide such exchanges. As a result, teachers and students tended to experience either a view of knowledge as *fait accompli*, or as idiosyncratic opinion that need no challenging.

In our earlier studies, and as with consistency scoring, we had set the bar high for categorizing respondents as criterialists. We looked for Cr scores at or above +2.7. The means here among all three groups of teachers remain below that threshold. However, they tended to rise slightly as the programs proceeded, and in the case of system C teachers, plateaued after year one. As we continually remind, the school culture’s vision of history education to an apparent degree sends messages that compete against efforts to enhance criterialist beliefs. Five days a week, 180 days a school year, these 45 teachers teach history surrounded by those messages. This may well mitigate what the TAH programs are able to achieve among the teachers and may explain why we had teachers tell us of their efforts to compartmentalize the two. TAH program participation is remunerated by small stipends but also is voluntary and therefore produces limited leverage on teachers’ daily work; whereas, the schooling culture is linked directly to the system and thus to teachers’ base bi-weekly paychecks, promotion, and salary advancement.
To further our understanding of the patterns we were seeing, we engaged in an analysis of all the teachers’ backgrounds (experience, U.S. history courses taken, history research methods courses taken, workshops attended, masters degree in history) as they entered the programs relative to the ways they scored on the BHQ (stance positioning, consistency scores, deltas). We wanted to know if those demographic factors mattered and if so to what degree they were linked and possibly predictive. In the case of experience, U.S. history courses taken, and workshops attended we calculated correlations and tested their significance. In the case of history research methods taken and masters degrees in history obtained, we compared means scores of teachers who had taken research method courses or had a masters in history and those who had not. Table 6 reports out the nature of these relationships, noting correlations and statistical significance, or approaching it, as relevant. In cases where the result is statistically non-significant, we observe NS as shorthand.

### Table 6. Teacher Background Demographics in Relation to Epistemic Beliefs (N = 45)

<table>
<thead>
<tr>
<th></th>
<th>Co Beliefs</th>
<th>Sub Beliefs</th>
<th>Cr Beliefs</th>
<th>Consistency Score Pre</th>
<th>Consistency Score Post</th>
<th>Consistency Score Delta</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total Years Taught (experience)*</td>
<td>–.15</td>
<td>–.07</td>
<td>–.08</td>
<td>.13</td>
<td>–.04</td>
<td>–.14</td>
</tr>
<tr>
<td></td>
<td>NS</td>
<td>NS</td>
<td>NS</td>
<td>NS</td>
<td>NS</td>
<td>NS</td>
</tr>
<tr>
<td>Total U.S. History Courses*</td>
<td>–.19</td>
<td>–.04</td>
<td>.34</td>
<td>.06</td>
<td>.01</td>
<td>–.04</td>
</tr>
<tr>
<td></td>
<td>NS</td>
<td>NS</td>
<td>(p = &lt;.05)</td>
<td>NS</td>
<td>NS</td>
<td>NS</td>
</tr>
<tr>
<td>Total History Workshops*</td>
<td>–.43</td>
<td>.17</td>
<td>.10</td>
<td>–.04</td>
<td>.04</td>
<td>.07</td>
</tr>
<tr>
<td></td>
<td>(p = ≤.01)</td>
<td>NS</td>
<td>NS</td>
<td>NS</td>
<td>NS</td>
<td>NS</td>
</tr>
<tr>
<td>No (n =23)</td>
<td>–.90</td>
<td>–.69</td>
<td>1.8</td>
<td>.79</td>
<td>.81</td>
<td>+.016</td>
</tr>
<tr>
<td>Yes (n = 22)</td>
<td>–1.6</td>
<td>–.73</td>
<td>1.9</td>
<td>.78</td>
<td>.82</td>
<td>+.038</td>
</tr>
<tr>
<td>Research Methods**</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No (n = 40)</td>
<td>–1.2</td>
<td>–.74</td>
<td>1.8</td>
<td>.79</td>
<td>.80</td>
<td>+.02</td>
</tr>
<tr>
<td>Yes (n = 5)</td>
<td>–1.5</td>
<td>–.44</td>
<td>2.3</td>
<td>.75</td>
<td>.86</td>
<td>+.12</td>
</tr>
<tr>
<td>Masters in History**</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No (n = 30)</td>
<td>–1.0</td>
<td>–.74</td>
<td>1.8</td>
<td>.79</td>
<td>.80</td>
<td>+.02</td>
</tr>
<tr>
<td>Yes (n = 6)</td>
<td>–1.5</td>
<td>–.44</td>
<td>2.3</td>
<td>.75</td>
<td>.86</td>
<td>+.12</td>
</tr>
</tbody>
</table>

* Pearson correlational analysis. **One-way ANOVA.
Although there are very few statistically significant results here to make generalizability claims with good confidence, some of the general patterns are nonetheless suggestive. The more history courses a teacher had taken, the greater the likelihood that their BHQ copier and subjectivist beliefs would be negatively correlated and their criterialist beliefs positively correlated. History workshop attendance reflected somewhat the same pattern, with the exception being the low positive correlation with subjectivist beliefs.

Similar directional patterns appeared with regard to whether teachers had taken history research methods courses (n=22) or not (n=23), and with respect to those with masters degrees in history (n=5). For those who had taken at least one research methods course in history, their copier and subjectivist beliefs were more negative on average and their criterialists beliefs more positive than those who had not taken such courses. For those with masters degrees, the same pattern emerged with the exception of subjectivist beliefs, which came somewhat closer to the point of ambivalence. Those with masters degrees also came to the end of the program (or after two years in the case of school system A teachers) more consistent in their beliefs (Consistency Score Post) and grew their epistemic consistency more so than their non-masters degree counterparts in the program (Consistency Score Delta).

To the extent that these patterns can be construed as meaningful (and we recommend some caution here), it appears that those with deeper initial historical knowledge gained from collegiate history course taking benefited most from efforts to grow and stabilize criterialist epistemic beliefs during the TAH programs. Generally, this would seem to make sense. Equipped with deeper historical knowledge and a stronger sense of what investigative expertise may look like provides epistemic advantages in a program that is designed to press participants toward acquiring criterialist beliefs. Gains, however, were relatively small. As such, they may again reflect the
countering the pressure of school history and school system culture—intentionally or not—to push participants to accept copier beliefs as *de rigueur*, thus attenuating programs’ efforts to shift epistemic beliefs.

**In the Classroom and Talks with Participants**

In our observations of teachers, we rarely saw a wholesale embrace of the investigative approach to understanding the past that the TAH programs attempted to immerse teachers. Rather, we witnessed frequent use of primary source documents that sometimes supplanted, but more often augmented the singular textbook narrative. That nation-state narrative dominated efforts. Coverage, as we have observed, appeared to be the guiding principle. Teachers occasionally taught students how to read and analyze the primary sources using historical thinking and reasoning strategies taught in the programs, but frequently these lessons became generic reading lessons in which teachers taught students strategies for finding main ideas and summarizing what was read. This was especially the case among elementary teacher participants and particularly in system C. In secondary schools, introducing primary sources became often a way of capturing student attention or an exercise in technical analysis, viewed as an end in itself and not as a means to come to a better interaction with the texts (documentary or pictorial).

There never seemed to be enough time to pursue analytic, historical thinking exercises with sources nor to stress the importance of pointing out that it is often through the analysis of accounts from the past that investigators come to make sense of that past. Rather, it was more common to see primary sources serve as curiosities—sometimes lurid or provocative—that were intended to spark student interest in a lesson otherwise driven by a lecture in Powerpoint format. School history goals of covering objectives and teaching the story of U.S. history as *fait accompli* dominated most classroom efforts. For a typical teacher participant, uses of primary sources in
lessons sometimes increased across observations as teachers responded to growth in their
application and use in the TAH programs, but the general structure of lessons did not change
dramatically.

Table 7 displays these relationships based on frequency counts obtained from the coding
categories we applied to our observation data. The frequency differentials reflect where teachers
invested their pedagogical energy.

**Table 7. Frequencies of Coding Categories in Classroom Observations**

<table>
<thead>
<tr>
<th>Coding Category</th>
<th>School System A (obs=16) Frequencies</th>
<th>School System B (obs=18) Frequencies</th>
<th>School System C (obs=28) Frequencies</th>
<th>Totals</th>
</tr>
</thead>
<tbody>
<tr>
<td>School Curricular Coverage Goals (e.g. objective stated)</td>
<td>10</td>
<td>9</td>
<td>9</td>
<td>28</td>
</tr>
<tr>
<td>Partial Exercise in Historical Thinking (e.g., use of primary source/s)</td>
<td>16</td>
<td>13</td>
<td>33</td>
<td>62</td>
</tr>
<tr>
<td>Full Exercise of Historical Investigation Practices (e.g., question-driven analysis of variety of accounts)</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>Not Engaged in Historical Thinking or Investigation during History Lesson (e.g., generic reading exercise, map exercises)</td>
<td>8</td>
<td>7</td>
<td>33</td>
<td>48</td>
</tr>
<tr>
<td>Total</td>
<td>35</td>
<td>31</td>
<td>75</td>
<td>141</td>
</tr>
</tbody>
</table>

At the risk of belaboring the point, it is important to keep in mind that we expressly asked
teachers to let us observe lessons that they believed reflected the goals of the TAH programs. That
so many of these lessons demonstrated only partial efforts to teach historical understanding as
rooted in historical thinking about the residue of the past likely resonates with the power of
coverage goals to challenge the full embrace of that approach. Some teachers may have come
away from the programs with the idea that increasing the number of primary source documents
used in their lessons was what the programs had in mind. Others may have thought that this was the best application they could tender given the fast-paced demands of covering the history curriculum. Teaching history as an investigative, interpretive enterprise can be time consuming because analyzing the past’s residue demands careful, systematic reasoning and subsequent debate over meaning. A history curriculum driven by coverage mandates designed to impart a broad swath of the nation-state’s narrative of development that appears in the *fait accompli* prose of the textbook and curriculum guide works at cross purposes with such an effort. Generally speaking, teachers facing the time-pressure dilemma appeared to be doing the best they could to manage it and this is what we observed. Notably, the approach that we observed here is not incompatible with the epistemic beliefs teachers tended to express while responding to the BHQ.

In our post-lesson conversations with teachers, this coverage agenda and the dilemmas it produced came up repeatedly. On the one hand, teachers talked at length about program goals as involving “teaching critical thinking skills, getting kids to analyze materials;” “moving away from the textbook;” “use of more sources;” “the idea of using primary sources in a more meaningful way;” “to let students investigate the past more;” “learning new strategies for teaching historical thinking;” and the like. On the other hand, they would as quickly remark, “But...there’s not enough time to have students question the documents because there’s too much to cover;” “I have started to simply use primary documents to cover what’s in the curriculum, not as a part of a historical investigation;” “even though the unit assessments don’t hold much weight in the school system, I see them as a constraining force for what I do in the classroom;” and “I feel very pressed to follow the system’s history curriculum and I don’t appreciate the historical inquiry approach because I really don’t see much application to my classroom.” Although we tallied only six direct
mentions of compartmentalizing out the approaches in program from school history, effectively a type of *de facto* compartmentalization seemed to going on for many teachers.

To the extent that full compartmentalization was not attractive to some teachers, their program application strategies—harking back to Ms. Oublier’s (see Cohen, 1990) partial adoption of a mathematics teaching reform program—hinged on taking bits and pieces of what they learned in the program and adapting them for use in their classrooms. Increasing use of primary source documents was a common strategy teachers discussed. But this use was frequently described in terms of “improving students’ reading abilities,” “using primary documents give teachers more techniques and skills to teach history,” “expanding our toolkits,” “providing useful reading strategies,” and “learning how to use one reading strategy per lesson.” The promise of introducing intriguing or juicy source materials to enhance general reading capacity was especially attractive to the elementary teachers and may reflect the powerful focus on the testing of reading and associated accountability requirements under NCLB. In part as such, it may have undermined the programs’ goal of inviting teachers to embrace the idea and pedagogy of historical-understanding-as-investigative-endavor because the teachers became more interested in using the documents for teaching tested reading skills than teaching historical understanding. Our observation data tended to bear that out.

Further, this partial appropriation of the tools and strategies offered by the programs seems in line with the epistemic portraits emerging from the BHQ data, because it does not require the kind of epistemic beliefs restructuring implied by a full embrace of the programs’ goals. In other words, it does not require teachers to question beliefs representative of the copier stance because documentary evidence, especially if used in isolation, can be treated as a doorway granting unmediated and authorless access to the past (or as the past itself). Conversely, as long as the
Residua of the past are seen as a repository of knowledge, and criteria for defending interpretations are lacking, eventual conflict among sources tends to be explained away as an issue of bias, sanctioning the view that the stories we decide to tell about the past are a mere matter of opinion.

Table 8 reflects a sense of the proportions of ideas that teachers choose to discuss in our conversations with them. TAH program goals were on their minds. When teachers discussed

Table 8. Frequencies of Interview Topics Discussed Per Coding Category

<table>
<thead>
<tr>
<th>Coding Category</th>
<th>School System A (int=16) Frequencies</th>
<th>School System B (int=18) Frequencies</th>
<th>School System C (int=28) Frequencies</th>
<th>Totals</th>
</tr>
</thead>
<tbody>
<tr>
<td>View of TAH Program Goals</td>
<td>1</td>
<td>15</td>
<td>4</td>
<td>20</td>
</tr>
<tr>
<td>Expressions of Compartmentalization</td>
<td>0</td>
<td>5</td>
<td>1</td>
<td>6</td>
</tr>
<tr>
<td>School History Coverage Goals and Assessments</td>
<td>4</td>
<td>5</td>
<td>4</td>
<td>13</td>
</tr>
<tr>
<td>Total</td>
<td>5</td>
<td>25</td>
<td>9</td>
<td>39</td>
</tr>
</tbody>
</table>

them, they were often compared or contrasted with school system goals and assessment practices. It was in that contrast where the adaptation principle became apparent in what they told us. The overall message appeared to be:

We like the TAH programs very much. We learn many good things about history and new ways of teaching in them. But only a portion of them (e.g., teaching reading skills, using primary sources to augment what we already do) are fully relevant to our history classrooms on a daily basis. The ones that are, we adapt to our circumstances and the demands those school culture contexts make on us because that’s who we are working for in the end. We think in terms of catchy teaching techniques and prize pedagogical activities that generate interest. We use them as primary criteria for determining how to draw from and adapt what we learn in the TAH programs.
The TAH program model made use of two types of professional developers, historians and what they called master teachers, who were drawn from the school systems themselves. The historians focused on deepening the teachers historical knowledge and immersed the teachers in investigative research activities so as to stress how to explore the past’s residua to better understand it and therefore model the process of historical investigation. The master teacher’s role involved helping the teachers transform program ideas into what they perceived to be adequate applications to history classroom contexts. In several cases, we observed that the view of history that the programs set out to foster got quite distorted in these translations.

Specifically, in observing and documenting the program interventions, we sometimes watched master teachers draw from the school systems’ technical orientations for thinking about teaching as they assisted the history teachers with these applications. As a result and without perceiveable irony, the teachers were encouraged by many of the school-based master teachers to take the complex ideas introduced around history as investigation and adapt it to the school systems’ coverage curriculum, assessment approaches, and technical orientations supporting it. Consequently, the observation and interview data we have described should not be construed as too surprising. It does reveal how the degree to which the school systems’ approach to thinking about and teaching history can function as a powerful over-riding influence on the TAH programs that were not fully aligned to it. It also reveals how difficult it is to effectively challenge the technical features of that system with this particular professional development model.

Discussion and Conclusion

As we have been alluding, history teachers in these three TAH programs appear to be squeezed in between two counterpoised modes of thinking and operating. Perhaps it can be useful to think of these differing modes as as constellations of attitudes about education—history
Pr

eparing Teachers

education in particular in this case. On the one hand, let us call what the teachers do as their principal paying occupation *school history*, and its constellation of attitudes as *technocratic education*. On the other hand, let us refer to the experience teachers have immersed in the TAH programs and the ways of thinking and operating there as *discipline-based history*, and its constellation of attitudes as the *education of judgment*. These serve as simple place markers that allow us to contrast the counter positionings of attitudes and therefore help account for the data we obtained from the teachers. We should also make clear that the two are not distinct. On our view, they hold some overlapping and shared attitudes. What distinguishes them is the value placed on those attitudes. One champions and stresses one attitude over others while the other inverts that valuation.

The technocratic education constellation of attitudes in school history values a clearly bounded and delineated curriculum, in this case, one that covers the story of national development. The story is long but linear. Nothing of importance—and almost everything accomplished by national leaders in the story arc is important—can or should be left out. Because of the story’s length, delivery speed and efficiency of cadence are crucial. The belief appears to be that if teachers tell the story, the students will get it. At least that appears to be the most desirable process because it seems most efficient.

Occasionally, students may be called upon to evaluate the right-ness of past actions and express their opinions in what is usually cast as an exercise in critical thinking. Such move does not usually address the nature of the narrative itself, nor compels teachers to provide students with criteria for investigating the past and rules of sound historical argument. In other words, students’ opinions are left unchallenged as long as students master the ostensible facts. Overall, there is much to do, much pressure to get it done, and so therefore there is little time to dawdle. To
facilitate this fast-paced process, standardization helps considerably. Variations of whatever sort threaten efficiencies and must not be tolerated for long. Much effort then is expended on reducing variance, in curriculum, in teaching practices, in student achievement and development, and so on (thus investments in standardized, high-stakes tests to identify and hopefully eliminate sources of variation, particularly in teaching practice).

Standardizing technique can help accomplish this end perhaps better than most other approaches, and so consequently, as an attitude, it is highly prized. Techniques with catchy acronyms that promise efficiencies and reduce variant outcomes are highly sought after, thus the acronym stew that tended to bubble up repeatedly when TAH program ideas met school history and the technocratic education attitudes in school systems. Professional judgment (e.g., how to vary approaches, ideas, or techniques to accommodate variation, say, in children’s development) tends to be construed as a necessary evil, but must be carefully constrained often at all costs. Exercises of teacher professional judgment, for instance, can increase variation to fevered pitches if left unchecked. Standardized curricula and teaching practices, pacing guides, acronym-laced standardizeable techniques, system-wide unit tests (as product quality control), and the like guard against runaway variation. Means that serve those ends often become fused with and inseparable from them. The teachers in our study knew this and appeared to have both mastered the understanding and appropriated it to their daily work lives (Wertsch, 1998) more or less without question, but not always without some almost inchoate forms of frustration.

Understanding the past and history to be one and the same, and viewing history through the prism of a fait accompli nation-state narrative, are powerfully consonant with these attitudes. Working from such understandings diminishes complexity, flattens out variation, reduces subjectivity to opinion so it is easily dismissed, fosters embrace of efficiencies, and permits
standardization techniques. It also permits easy expenditures on standardized testing that appear to provide assurances of quality control. It is all deeply seductive and epistemically reassuring.

Discipline-based history and the constellation of attitudes surrounding judgment education share features in common with the former. For example, discipline-based history also values technique. However, how it values it and to what end tends to be inverted with respect to school history. Technique serves as a bit player in the larger scheme of cultivating judgment that results in enhanced learning. Almost anyone can master a technique without necessarily knowing when it is best applied, or rely on a different, perhaps more appropriate one. Judgment recognizes this problem and seeks to cultivate criteria to bend technique to its need. As an attitude, it is much more tolerant of variation as long as that variation can be dealt with and defended by judgment criteria. Therefore, it invests far more heavily in teaching judgment and views strength in variation.

Pressing technique (e.g., assessing source status, utilizing organizing concepts, reading and writing) into the service of judgment in disciplinary history is consonant with a different view and attitude about the past. Here, the past is understood as eminently unstable and largely unknowable in any complete sense. It is messy, complex, prone to peculiarities of historical context and temporal anchor and tends to trouble investigators’ reason. It invites but limits simultaneously. Variation abounds and resists reduction. Working among such difficulties and challenges requires a knower who holds powerful judgment criteria for regulating techniques that allow her access to an understanding of the past. Under these circumstances, techniques for knowing defy easy acronymal classification. They cannot be readily standardized. For school history, such a view and its accompanying attitudes border on the nightmarish.
School history welcomes and courts copier epistemic beliefs because they are a much better fit with their constellation of technocratic attitudes. Disciplinary history is much more comfortable with criterialist epistemic beliefs because it works from and prizes professional judgment. Copier beliefs in an uncertain world allow that world to be more easily managed and controlled. They promise efficiencies of navigation and simplicities that criterialist beliefs cannot recommend. Criterialist beliefs are also more difficult to cultivate. They require contexts in which to cautiously practice working them out, and many of them. School systems on their own do not appear willing to invest such time in their teachers, especially given the testing and accountability climate imposed on them. TAH programs such as the three we described here are valuable in the sense that they give teachers time to practice cultivating criterialist beliefs (e.g., through the historical research investigations they undertook).

Despite these opportunities, the return to school history and the technocratic education attitudes, to which teachers were already most acclimated, attenuates what was learned, and as a result much of the TAH program interventions were at least partially undone. The epistemic tensions that these experiences created for teachers must have created some consternation. Compartmentalization likely functioned for some as a reasonable psychic defense against the rather schizophrenic juxtaposition of attitudes they witnessed. Developing consistent epistemic beliefs was likely another casualty. The TAH program attitude constellations would tug and press in one direction while daily school history demands pulled in another. Mastery of ideas and understandings encountered in the TAH interventions may have been sampled by the BHQ and picked up in teacher talk in interviews (Wertsch, 1998). But the lack of full appropriation of those ideas and understandings was also sampled, again by the BHQ in terms of epistemic consistency.

---

2 Richard Brown (1966, 1996) remarked on similar outcomes on many of the professional development ventures the Amherst Project engaged in over 40 years ago.
scores, but more clearly in the observational data surrounding practices and in watching master teachers encourage participants to transform program ideas back into school history effects.

It is probably fair to ask, given our opening question about what people are getting for their investments in this TAH professional development program, whether it appears worth it to subject history teachers to such mixed messages and counterpoised attitudes, especially since the intended outcome is so compromised. One way to respond to this query is to observe a subsequent development in one of the three school systems that came in the wake of us reporting out these data and their attendant analyses.

Upon listening to the explanatory theory outlined here, directors in school system B elected to write another TAH grant proposal on the heels of the one reported on here. It was later funded and now is in operation. In this latest iteration, system B made a bold move to allow modifications to the U.S. history curriculum despite pressure from the state to make sure that curriculum aligned with the mile-wide state standards. For participating teachers, they gave express privilege with school board approval to permit them to engage in full-blown historical investigations (constructed on disciplinary-history attitudes) around topics/events of their choosing, even if it meant that pressing them into practice resulted in their inability to cover the entire mandated nation-state narrative. Teachers selected investigative episodes and constructed them in the context of work with historians on the new TAH grant program. They then taught them to their students. We are currently engaged in observations of these investigative units.

Directors then commissioned our evaluation team to help a small group of teachers write student assessments aligned with these investigations. The goal is to use these assessments to gauge growth in historical thinking and understanding among the students taught by the TAH program teachers. To test the value of endorsing this approach on historical understanding among
students, a comparison group matched as closely as possible to the students in the investigation classrooms will also take the assessments. Results will be analyzed and compared. The directors in school system B hope, of course, that the investigative approaches and the results they produce provide evidence that historical understanding can be deepened by such an approach, that in-depth, careful, slow-paced historical study trumps fast-paced curricular coverage.

The jury on this latest effort is still out. However, this sort of change, one that runs against the entrenched and, some would argue, uninspiring technocratic attitudes of schooling and school history, was made possible largely by the wake of evaluation reports from previous “less than fully successful” TAH programs. That such data and the reports emanating from them can evoke these sorts of experimentations at least suggests that research-rationalized public investments in growing teachers’ historical knowledge hold promise for potentially deepening student understanding. In the end, that has been the justification for these investments all along.
References


Appendix A

Beliefs about History Questionnaire List of Statements

1. It is fundamental that students are taught to support their reasoning with evidence.
2. History is simply a matter of interpretation.
3. A historical account is the product of a disciplined method of inquiry.
4. Students who read many history books learn that the past is what the historian makes it to be.
5. Disagreement about the same event in the past is always due to lack of evidence.
6. Good students know that history is basically a matter of opinion.
7. Students need to be taught to deal with conflicting evidence.
8. Historical claims cannot be justified, since they are simply a matter of interpretation.
9. Good general reading and comprehension skills are enough to learn history well.
10. Since there is no way to know what really happened in the past, students can believe whatever story they choose.
11. History is a critical inquiry about the past.
12. The past is what the historian makes it to be.
13. Comparing sources and understanding author perspective are essential components of the process of learning history.
14. It is impossible to know anything for sure about the past, since no one of us was there.
15. Knowledge of the historical method is fundamental for historians and students alike.
16. The facts speak for themselves.
17. Students need to be aware that history is essentially a matter of interpretation.
18. Reasonable accounts can be constructed even in the presence of conflicting evidence.
19. Even eyewitnesses do not always agree with each other, so there is no way to know what happened.
20. Teachers should not question students’ historical opinions, only check that they know the facts.
21. History is the reasonable reconstruction of past occurrences based on the available evidence.
22. There is no evidence in history.