

- McCall, L. (2005). The complexity of intersectionality. *Signs*, 30(3), 1771–1800.
- Pandel, H.-J. (1987). Dimensionen des Geschichtsbewußtseins. Ein Sersuch, seine Struktur für Empirie und Pragmatik diskutierbar zu machen. *Geschichtsdidaktik*, 12, 130–142.
- Rüsen, J. (1993). *Studies in Metahistory*. Pretoria, South Africa: Human Sciences Research Council.
- Rüsen, J., Fröhlich, K., Horstkötter, H., & Schmidt, H.-G. (1991). Untersuchungen zum Geschichtsbewußtsein von Abiturienten im Ruhrgebiet. In B. v. Borries, H.-J. Pandel, & J. Rüsen (Eds.), *Geschichtsbewußtsein empirisch* (pp. 221–344). Pfaffenweiler, Germany: Centaurus.
- Sauer, M. (2006). Kompetenzen für den Geschichtsunterricht—ein pragmatisches Modell als Basis für die Bildungsstandards des Verbandes der Geschichtslehrer. *Informationen für den Geschichts- und Gemeinschaftskundelehrer*, 74, 7–20.
- Sauer, M. (2009). *Geschichte unterrichten. Eine Einführung in die Didaktik und Methodik*. Kallmeyer, Germany: Seelze.
- Schönemann, B. (2012). Geschichtsbewusstsein—Theorie. In M. Barricelli & M. Lücke (Eds.), *Handbuch Praxis des Geschichtsunterrichts, Bd. 1* (pp. 98–111). Schwalbach/Ts., Germany: Wochenschau-Verlag.
- Schörken, R. (1972). Geschichtsdidaktik und Geschichtsbewußtsein. *Geschichte in Wissenschaft und Unterricht*, 23, 81–89.
- Schott, R. (1968). Das Geschichtsbewußtsein schriftloser Völker. *Archiv für Begriffsgeschichte*, 12, 166–205.
- Schreiber, W. (2008). Ein Kompetenz-Strukturmodell historischen Denkens. *Zeitschrift für Pädagogik*, 54(2), 198–212.
- Seixas, P. (2005). Historical Consciousness: A Scheme of Progress in Knowledge for a Post-Progressive Age. In J. Straub (Ed.), *Narration, Identity, and Historical Consciousness: The Psychological Construction of Time and History* (pp. 44–98). New York: Berghan.
- Sonntag, K. (1932). *Das geschichtliche Bewusstsein des Schülers. Ein Beitrag zur Bildungspsychologie*. Erfurt, Germany: Stenger.
- Straub, J. (Ed.). (2005a). *Narration, Identity, and Historical Consciousness*. New York: Berghan.
- Straub, J. (2005b). Telling Stories, Making History: Toward a Narrative Psychology of the Historical Construction of Meaning. In J. Straub (Ed.), *Narration, Identity, and Historical Consciousness* (pp. 44–98). New York: Berghan.
- Verband der Geschichtslehrer Deutschlands (VGD) (Eds.). (2006). *Bildungsstandards Geschichte. Rahmenmodell Gymnasium 5.-10. Jahrgang*. Schwalbach/Ts., Germany: Wochenschau-Verlag.
- Verband der Geschichtslehrer Deutschlands (VGD) (Eds.). (2010). *Bildungsstandards Geschichte (Sekundarstufe I)*. <http://www.geschichtslehrerforum.de/Standards/index.html>.

2

THE DIFFICULTY OF ASSESSING DISCIPLINARY HISTORICAL READING

Abby Reisman

Historical thinking depends on the ability to reason about written text. Writing in 1899, American historian Frederick Jackson Turner encouraged history educators to replace “the old ideal of history as pleasant literature” with “the ideal of history as a discipline of the mind, valuable particularly as a training of judgment in the criticism of material like that which is placed before the citizen in current political and industrial questions” (p. 301). Nearly a century later, the Bradley Commission on History in Schools—a body that advocated a renewed emphasis on historical study in K-12 social studies instruction—recommended “training in critical judgment based on evidence, including original sources” (Bradley Commission, 1989, p. 23).

It seems somewhat perplexing to consider, then, that history assessments have not mirrored this consistent emphasis on the value of reasoning across sources. By 2011, twenty-six of the United States required testing in history/social studies, with half of these using multiple-choice-only measures and the other half using a combination of constructed-response and multiple-choice (Martin, Maldonado, Schneider, & Smith, 2011). Fogo’s (in press) account of California’s 1987 *History-Social Science Framework* shows how broad curricular recommendations became discrete content standards that were ultimately assessed with decontextualized multiple-choice items on the (now defunct) Standardized Testing and Reporting (STaR) exams. And Reich (2009) discovered a striking misalignment between the reasoning of students who achieved correct answers on the 10th grade New York Regents exam in global history and the historical analysis standard that the test claimed to measure. When we look to explain this persistent misalignment between the stated goals of history instruction and the tests designed to assess them, we find that test-makers have been more concerned with efficiency and psychometric reliability than with disciplinary validity (Wineburg, 2004). But I

would like to make the case that a more nagging problem underlies the longevity of multiple-choice tests as historical assessments: the difficulty of disentangling disciplinary historical reading—or, the ability to evaluate the reliability of sources in order to construct an intertextual account of the past—from students' incoming background knowledge, on the one hand, and general reading comprehension, on the other.

Historical Reading and the Common Core Standards

As has been well documented (cf. Lee, 2005), historical thinking depends upon, but extends beyond historical reading. In the United States, the few efforts to design assessments that move beyond factual recall to measure historical thinking have primarily focused on historical reading, students' ability to construct an account of the past from written sources. Wineburg (1991a, 1991b) characterized disciplinary historical reading as an epistemological orientation towards texts that allows the historian to view the texts as human constructions whose probity can and should be interrogated. Wineburg further distilled three discrete heuristics that historians applied while reading historical texts: sourcing (considering the document's source and purpose), contextualization (placing the document in a temporal and spatial context), and corroboration (comparing the accounts of multiple sources against each other).

While neither Wineburg's work nor the subsequent research on disciplinary historical reading was initially framed within a broader reading research agenda, the work dovetailed well with other developments in American education. In 2002, the Rand Corporation, an American non-profit policy think tank, published a report on reading comprehension that highlighted the dearth of research on reading comprehension in middle and high school, even as demands for advanced literacy continued to grow (Snow, 2002). Soon after, the Carnegie Corporation published *Reading Next* in 2004, which charted "an immediate route to improving adolescent literacy" (Biancarosa & Snow, 2006). The report was followed by the formation of the Carnegie Council for Advancing Adolescent Literacy and the subsequent publication of several additional reports, culminating with *Time to Act* in 2010, which "pinpoints adolescent literacy as a cornerstone of the current education reform movement."

By emphasizing domain-specific literacy and the particular demands of content area texts, the reports provided a national platform for work on historical reading. Two findings, in particular, related directly to history instruction: the first was that students need exposure to a range of textual genres in order to be prepared for college; the second was that how one reads differs by content area (Biancarosa & Snow, 2006; Carnegie Council on Advancing Adolescent Literacy, 2010; Heller & Greenleaf, 2007; National Institute for Literacy, 2007). In other words, historians read and ask different questions of texts than do scientists or poets; primary sources, presumably, afford students the opportunity to practice these domain-specific reading practices. These two findings

have most recently found their way into the Common Core State Standards for English Language Arts and Literacy in History/Social Studies, Science, and Technical Subjects (2010), literacy standards that have been adopted by all but seven states. The first appears as a “key design consideration,” with the authors citing “extensive research establishing the need for college and career ready students to be proficient in reading complex informational text independently in a variety of content areas” (p. 4). The second appears later in the document: in the introduction to reading in the remaining subjects—history, science, and technical subjects—the authors state that “college and career ready reading in these fields requires an appreciation of the norms and conventions of each discipline” (p. 60).

The energy and momentum accompanying the Common Core Initiative has been, on the one hand, exhilarating. The initiative promises a national platform for assessments on *historical reading*, a reality few would have considered possible a mere decade ago. On the other hand, the fact that the standards emerged from the work on adolescent literacy, with only cursory engagement with the literature on historical thinking and reading, has direct implications for the sorts of assessments that will likely emerge. A close look at the Reading Standards for History/Social Studies (p. 61) finds that they are almost indistinguishable from the Reading Standard for Informational Texts (pp. 39–40), aside from the inclusion of inter-textual reading as a core feature of the history/social studies standards. Conspicuously absent from Reading Standards for History/Social Studies, however, is any reference to history or historical knowledge. This stands in contrast to the consensus that has emerged in the research on historical thinking: that meaningful historical thinking requires familiarity and facility with disciplinary ways of interpreting historical texts, an appreciation of the interpretive nature of historical knowledge, and the application of conceptual, narrative, and discrete factual knowledge. If test developers infer from these standards that historical reading assessments can be designed without considering students’ background knowledge, they will fall into an old trap.

Role of Background Knowledge in Historical Reading

“Disciplinary literacy” has become something of a buzzword in educational circles in the United States, thanks in large part to the Common Core State Standards. In the name of promoting “disciplinary literacy,” educators have focused on identifying reading “strategies” employed by expert historians, or pinpointing the particular linguistic demands of history textbooks. The intention—to focus attention on *reasoning* and *interpretation* rather than on memorization and retention—should be applauded. However, it is also important to recognize the essential role that background knowledge plays in allowing experts to employ these various disciplinary reading strategies. A baseline level of knowledge is evident even in the research that has been marshaled most frequently to make the case for teaching disciplinary historical reading. Wineburg (1991b, 1998)

finds repeatedly that deep expert knowledge is not essential for historians to navigate a particular problem in history—in his studies, Medievalists and Sinologists expertly read and contextualize texts about the Battle of Lexington and Abraham Lincoln. However, if we consider the context of typical classrooms, we still must ask: what baseline level of knowledge did these non-experts possess? Certainly, they were familiar with the basic narrative of the American Revolution. Certainly they possessed knowledge of Abraham Lincoln’s signature achievement—the Emancipation Proclamation.

In a more recent but comparable study, Baron (2012) examined how expert historians “read” a historical space, namely, the Old North Church in Boston, Massachusetts. Arguing that Wineburg’s heuristics for reading documentary texts did not account for how experts reason about space, Baron identified five new heuristics from the historians’ think-aloud protocols and argues that familiarity with these heuristics might allow novices to begin to ask “What are the multiple time periods evident in this building, and what do they tell me about its history?” (p. 844). Yet, Baron’s historians heavily depend on their historical knowledge and experience. Their “heuristics” represent different ways that they used their knowledge to make inferences about the building; indeed, three of the five heuristics—origination (considering the building’s origin), stratification (identifying the multiple strata of time evident in the building), and empathetic insight (considering the affective response of people who occupied the space at particular historical moments)—were largely unavailable to those historians with little background knowledge.

Van Boxtel and Van Drie (2012) asked a different question related to historical texts. What allows students to successfully contextualize historical images and documents: historical knowledge or strategy use? The authors found that providing students with knowledge of key substantive historical concepts, and helping them construct an associative network around those concepts, was most predictive of student success on contextualization tasks. This study underscores the importance of historical knowledge in leveraging student reasoning with and about historical texts.

The entwined nature of historical knowledge and historical reading poses formidable challenges to large-scale history assessment developers. Below I discuss two examples of history assessments where the effort to capture historical reading was caught between the confounding factors of background knowledge and reading comprehension.

Example 1: Learning-Based Assessments of Historical Understanding

In a 1994 special issue of *Educational Psychologist*, Eva Baker describes a six-year effort to design performance-based assessments of historical understanding (Baker, 1994). The project, a collaboration between UCLA’s National Center for Research on Evaluation, Standards, and Student Testing (CRESST) and Los Angeles Unified

School District, was an effort to create large-scale assessments that would capture broad patterns in student achievement or educational reform, and simultaneously provide classroom teachers important formative information about their students' learning and content understanding. An important distinction must be drawn between “historical thinking”—or disciplinary ways of thinking about the past—and “historical understanding.” Baker’s project did not purport to measure “historical thinking”—a term that had yet to be popularized; indeed, Baker’s conception of “historical understanding” can best be understood as the *result* of historical thinking—namely, a textured understanding of what happened, say, during the Lincoln-Douglas debates (see Figure 2.1). Yet, in attempting to design an assessment of historical understanding that required students to reason across multiple documents, Baker’s efforts shed light on the potential pitfalls of measuring historical reading.

In designing the assessment, the researchers faced three constraints. First, they had to avoid designing assessments that captured stable traits or general talents that would not be affected by historical study. In other words, the assessment needed to measure historical understanding as something that was distinct from fluid writing, for example. Second, to ensure content validity, the assessments needed to invite student interpretation. Yet, they needed to simultaneously remain “relatively insensitive to varying content emphasis and epistemological differences among history experts and teachers” (p. 99). This constraint raised a natural dilemma: if students’ thinking processes are given greater weight than the specific substance of their answers (so as not to penalize unpopular interpretations), at what point does the assessment cease to measure historical understanding? The third constraint was more universal: any large-scale assessment is limited by practical scoring considerations and feasibility.

Writing Assignment

Imagine that it is 1858 and you are an educated citizen living in Illinois. Because you are interested in politics and always keep yourself well informed, you make a special trip to hear Abraham Lincoln and Stephen Douglas debating during their campaigns for the Senate seat representing Illinois. After the debates you return home, where your cousin asks you about some of the problems that are facing the nation at this time.

Write an essay in which you explain the most important ideas and issues your cousin should understand. Your essay should be based on two major sources: (1) the general concepts and specific facts you know about American History, and especially what you know about the history of the Civil War; (2) what you have learned from the readings yesterday. Be sure to show the relationships among your ideas and facts.

FIGURE 2.1 CRESST performance assessment in history writing prompt (Baker et al., 1992, p. 11)

The assessments ultimately took the form of explanation tasks whereby students read two primary sources that offered competing positions on a historical topic, and explained to a friend or colleague what was happening and why it was important (see Figure 2.1). Developers decided to incorporate primary sources not only for reasons of disciplinary validity—historians, of course, read primary sources—but also to “level the field” for students who may have received varying exposure to the historical topic. The assessment was designed to align with extant processing models that viewed text-comprehension, integration, and application as the core features of “deep content understanding” (Glaser, 1992). The assumption was that students would draw both on their background knowledge and understanding of the historical topic as well as on the information in the provided texts to write the essay.

Six years of field-testing yielded mixed results that informed subsequent design decisions. To begin, researchers quickly discovered that students rarely, if ever, incorporated outside knowledge—even those students whose teachers claimed to have taught the content. Moreover, students generally did not evaluate or assess historical significance, but rather “tried to cover all information with an equally light hand” and they made numerous factual and conceptual errors (p. 101). This finding stood in sharp contrast to the essays written by experts in their subsequent expert-novice study, who incorporated extensive prior knowledge and organized their essays around broad principles, using the provided texts to illustrate key points. In subsequent iterations, developers incorporated a 20-item, short-answer prior knowledge measure that preceded the main task. The goal of this measure was both to activate student prior knowledge and to assess the relationship between such knowledge and student performance on the written component of the assessment. The researchers also added a 14–15-item multiple-choice literal comprehension test to determine if students could make literal sense of the historical texts that were provided. Although Baker did not report on the relationship between these measurement components, their inclusion speaks to the difficulty of disentangling historical reading from background knowledge and literal reading comprehension.

Baker and her colleagues also found disappointing patterns among raters. Initially four expert teachers identified essential criteria for historical understanding (e.g., evidence of historical analysis, detail, etc.), yet factor analyses conducted on these criteria yielded a single factor; in other words, “rating only one element (e.g., logical structure) or using a single overall rating would provide the same amount of information as scoring multiple elements” (p. 100). Ultimately a scoring rubric was designed that included the following dimensions: (1) Overall content quality; (2) Prior knowledge; (3) Principles/themes; (4) Text detail; (5) Misconceptions; and (6) Argumentation. Later, when factor analyses were conducted on these dimensions, researchers found a consistent two-factor solution, in which Prior Knowledge, Principles, and Overall Content Quality loaded on one factor, while Text Detail and Misconceptions formed another. The factor

structure matched differences between expert and novice essays. Subsequent exploratory analysis found an interaction between raters' content knowledge and their scores: "Raters with less knowledge tended to overvalue text material in students' writing, and consequently the relation between their ratings of text detail and rater judgment of general content quality increased" (p. 103). Raters' content knowledge also had implications for what was considered a "misconception."

CRESST has engaged in several assessment development projects since the 1990s—most recently in response to the content literacy demands of the Common Core State Standards. However, this earlier effort stands out for both its duration and its attempt to substantively engage with the demands of historical subject matter. For those same reasons, it demonstrates the inherent challenges in designing assessments of historical understanding based on primary sources, assessments that demand that students bring a degree of background knowledge and basic reading comprehension. One could argue that the assessment that was ultimately developed—an explanation task based on two documents—does not constitute authentic engagement with primary sources. Because the assessment ultimately equated historical knowledge with the ability to summarize the views expressed in primary sources, with no opportunity for argumentation or interpretation, one could argue that it demanded little more from students than general reading comprehension. In the example below, I demonstrate how the effort to design an argumentation task around multiple sources encountered similar challenges.

Example 2: Advanced Placement Document-Based Question

The Advanced Placement program was developed in the 1950s in an effort to give "able school boys and girls" an opportunity to challenge themselves with advanced coursework (Schneider, 2009). Run by the College Board since 1955, college credit for the course is determined by one's score on a summative exam. The signature feature of the history exams is the Document-Based Question (DBQ), first created in the 1970s in an effort to prompt students to do more than simply recite memorized factual information—their typical response to the non-document essay prompt. The DBQ, by contrast, assessed students' ability to assess and synthesize multiple primary sources. To many, the AP's DBQ represents the gold standard in authentic historical assessment, a model for assessing disciplinary literacy. Others have criticized the DBQ for being too broad and unfocused to provide classroom teachers with formative information about their students' learning (Breakstone, Smith, & Wineburg, 2013). A peek into the development of the current DBQ sheds some light on its shortcomings as an assessment of historical reading.

Writing in the American Historical Association's publication *Perspectives* in 1983, Stephen Klein, an Educational Testing Services (ETS) consultant to the College Board Development Committee for Advanced Placement American History, discussed the committee's reasoning for shortening the DBQ on the

American History examination from 15–20 documents to a mere half dozen. The committee had discovered two persistent problems in student responses to the longer DBQ: (1) the extensive, self-contained document set provided students with sufficient information about the given topic that they did not need—and, indeed, rarely bothered—to incorporate prior background knowledge into their written responses; and (2) because the prompts emphasized the *analysis* and *synthesis* of documents, it was less apparent how students were to demonstrate their ability to assess relevance and reliability, skills that arguably constitute the heart of historical reading. Klein shows how both skills were gradually deemphasized in the essay prompt: by 1975, both requirements—to incorporate background information not included in the documents and to assess the reliability of the evidence—became optional.

Although the decision to shorten the DBQ initially emerged from practical considerations, not least of which was the time required to assemble longer document sets, the hope was that it would address the two concerns discussed above. The committee reasoned that a shorter document set would force students to incorporate outside knowledge, though it might inadvertently result in the inverse problem: that students would use the documents as mere launching points for a recitation of memorized factual information. The problem of assessing reliability was even more elusive. As Klein (1983) explained,

Relevance and reliability are determined either by a pat formula having nothing to do with historical knowledge, or by a depth of historical knowledge unlikely to be possessed by a survey course student. An example of the former would be: “A witness to an event is more reliable than someone who hears about it second-hand.” An example of the latter would be: “Calhoun was a less reliable judge of Jackson’s political motives after their falling out in 1830 than before.” The nice thing about the latter example is that one could argue just the opposite . . . but the main point is that whichever side one argues for, substantial historical knowledge would be necessary.

(p. 23)

The committee concluded that in order to keep students focused on the actual documents, the prompt would have to ask about historical significance. As Klein explained, “The committee assumed that, ultimately, questions about historical significance would probably reduce themselves to “To what extent do the documents support and/or contradict what you already know about historical topic X?”” (p. 24). Furthermore, as Baker and her colleagues concluded, the topic would have to be a mainstream topic that most students could be expected to have encountered in their course.

In Fall 2014, the College Board released a new curricular framework for the AP United States History Exam (College Board, 2014). The DBQ remains a key feature of the free-response section, though the number of documents included

in the prompt has been reduced to 7 after several decades of including 9 or 10 documents. The rubric reflects more significant changes, stipulating that document analysis should address “intended audience, purpose, historical context, and/or point of view” and that students should support their arguments with “analysis of historical examples outside the documents.” It remains unclear whether and how students will respond to these changes. Klein found in the administration of the first “shorter” DBQ in 1982 that few students incorporated outside knowledge. If the document sets do contain sufficient information, and students are not required to evaluate the reliability or relevance of the evidence, to what extent is the exam assessing general reading comprehension rather than disciplinary reading and historical understanding?

Conclusion

The recent focus on disciplinary historical reading in the United States brings with it the promise of long-awaited assessment reform, a shift away from multiple-choice questions about decontextualized facts towards assessments that ask students to interpret and reason across multiple historical texts. However, as is evident both in the research literature and from previous efforts to design large-scale assessments of historical reading, reasoning about texts requires a degree of background knowledge. Furthermore, any large-scale assessment of historical reading must address the varying degrees of familiarity with a given topic that students will bring to the task. If the assessment includes background knowledge (to “level the field”), and does not explicitly require students to assess reliability and relevance, at what point does it become a reading test that simply asks students to summarize or explain the substance of the historical documents? Assessment developers will need to think carefully about what knowledge students are expected to bring to the task, how students will be asked to use this knowledge when engaging with texts, and whether or not any background knowledge will be provided. Without attending to the relationship between background knowledge and historical reading, new, innovative assessments that attempt to address the historical reading standards of the Common Core run the risk of foundering on the same challenges that have limited previous assessment initiatives.

Lest I be accused of ending on a dire note, I should add the silver lining: the above conundrum concerns large-scale assessment-designers, not individual classroom teachers. Whereas the large-scale designer cannot know what background knowledge students bring to the test, the classroom teacher can know whether students have sufficient background knowledge to interpret certain texts. A teacher who is interested in assessing historical reading might think of the process as three-fold: (1) assessing students’ background knowledge; (2) assessing students’ ability to evaluate the reliability and relevance of textual evidence; and (3) assessing students’ ability to synthesize multiple documents into an account. These three assessments need not be administered together. The assessment of

(2), students' ability to evaluate a source's reliability and relevance, for example, could be comprised of intermittent short assessments, such as the ones designed by the Stanford History Education group (beyondthebubble.stanford.edu). By contrast, an assessment of (3), students' ability to synthesize documents, might be designed as a short DBQ about a topic/event with which students are unfamiliar but that occurred in a context and time period that they have studied extensively. The goal is for teachers to disentangle the various threads that comprise historical thinking in order to better target instruction and assessment.

References

- Baker, E. (1994). Learning based assessments of historical understanding. *Educational Psychologist*, 29(2), 97–106.
- Baker, E., Aschbacher, P., Niemi, D., & Sato, E. (1992). *CRESST performance assessment models: Assessing content area explanations*. Los Angeles, CA: University of California, National Center for Research on Evaluation, Standards, and Student Testing.
- Baron, C. (2012). Understanding historical thinking at historic sites. *Journal of Educational Psychology*, 104(3), 833–847.
- Biancarosa, C., & Snow, C. (2006). *Reading next—A vision for action and research in middle and high school literacy: A report to Carnegie Corporation of New York* (2nd ed.). Washington, DC: Alliance for Excellent Education.
- Bradley Commission on History in Schools. (1989). Building a history curriculum: Guidelines for teaching history in schools. In P. Gagnon (Ed.), *Historical literacy: The case for history in American education* (pp. 16–47). Boston, MA: Houghton Mifflin.
- Breakstone, J., Smith, M., & Wineburg, S. (2013). Beyond the bubble in history/social studies assessments. *Phi Delta Kappan*, 94(5), 53–57.
- Carnegie Council on Advancing Adolescent Literacy. (2010). *Time to act: An agenda for advancing adolescent literacy for college and career success*. New York, NY: Carnegie Corporation of New York.
- College Board. (2014). The AP United States History Exam. Retrieved from http://apcentral.collegeboard.com/apc/members/exam/exam_information/2089.html
- Common Core State Standards for English Language Arts & Literacy in History/Social Studies, Science, and Technical Subjects*. National Governors Association Center for Best Practices, Council of Chief State School Officers, Washington D.C. Retrieved from http://www.corestandards.org/wp-content/uploads/ELA_Standards.pdf
- Fogo, B. (in press). The making of California's history standards: Enduring decisions and unresolved issues. *The History Teacher*.
- Glaser, R. (1992). Learning, cognition, and education: Then and now. In H. L. Pick, Jr., P. van den Broek, & D. C. Knill (Eds.), *Cognition: conceptual and methodological issues* (pp. 239–265). Washington, DC: American Psychological Association.
- Heller, R., & Greenleaf, C. (2007). *Literacy instruction in the content areas: Getting to the core of middle and high school improvement*. Washington, DC: Alliance for Excellent Education.
- Klein, S. (1983). The genesis of shorter document-based questions in the advanced placement American history examination. *American Historical Association Perspectives* (May/June): 22–24.
- Lee, P. (2005). Putting principles into practice: Understanding history. In M. S. Donovan and J. D. Bransford (Eds.), *How students learn: history in the classroom* (pp. 31–74). Washington, DC: National Academies Press.

- Martin, D., Maldonado, S., Schneider, J., & Smith, M. (2011). A report on the state of history education: state policies and national programs. *National History Education Clearinghouse*. Retrieved from http://teachinghistory.org/system/files/teachinghistory_special_report_2011.pdf
- National Institute for Literacy. (2007). *What content-area teachers should know about adolescent literacy*. Washington, DC: Author.
- Reich, G. (2009). Testing historical knowledge: standards, multiple-choice questions and student reasoning. *Theory and Research in Social Education*, 37(3), 325–360.
- Schneider, J. (2009). Privilege, equity, and the advanced placement program: tug of war. *Journal of Curriculum Studies*, 41(6), 813–831.
- Snow, C. (2002). *Reading for understanding: toward a research and development program in reading comprehension*. Santa Monica, CA: RAND.
- Turner, F. J. (1899, October). *Educational Review*, 18, 301–304.
- Van Boxtel, C., & Van Drie, J. (2012). “That’s in the time of the Romans!” Knowledge and strategies students use to contextualize historical images and documents. *Cognition and Instruction*, 30(2), 113–145.
- Wineburg, S. (1991a). Historical problem solving: A study of the cognitive processes used in the evaluation of documentary and pictorial evidence. *Journal of Educational Psychology*, 83(1), 73–87.
- Wineburg, S. (1991b). On the reading of historical texts: Notes on the breach between school and academy. *American Educational Research Journal*, 28(3), 495–519.
- Wineburg, S. (1998). Reading Abraham Lincoln: An expert/expert study in the interpretation of historical texts. *Cognitive Science*, 22(3), 319–346.
- Wineburg, S. (2004). Crazy for history. *Journal of American History*, 90(4), 1401–1414.

3

HERITAGE AS A RESOURCE FOR ENHANCING AND ASSESSING HISTORICAL THINKING

Reflections from the Netherlands

Carla van Boxtel, Maria Grever, and Stephan Klein

Introduction

In every country, there are historical issues that are vital in collective memory and repeatedly give rise to public debates. In the Netherlands, the transatlantic slave trade and its associated traces of the past aptly illustrate the sensitivity of such issues. It is only very recently and hesitantly that the Dutch government has acknowledged the historical role of the Dutch. For instance, in 2002 a national slavery monument was unveiled in Amsterdam and the annual commemoration of the Dutch abolition of slavery on July 1 implemented. It is only in the last ten to fifteen years that the topic has been integrated in both academic historiography and school history curricula, although specialists are still very critical. With respect to school history, they argue that the slave trade is often represented as a side story and that the emphasis is mainly on the abolition by the Dutch, ignoring the agency of enslaved people themselves (Van Stipriaan, 2007).

Whereas the Dutch involvement in slavery has been acknowledged at the national level, for many descendants it remains an emotionally charged issue. Part of the Afro-Caribbean Dutch community has demanded substantial “reparations” for what they call the “Black Holocaust”. The recent controversy in the Netherlands about the phenomenon “Black Pete”, which attracted international attention, can also be connected to the legacy of Dutch slavery. Every year in November, Dutch children eagerly look forward to the arrival of St. Nicholas and his Black Petes (Zwarte Pieten), coming from Spain on a steamboat with lots of presents. However, particularly since the 1980s with the arrival of migrants from the Dutch former colonies Suriname and the Antilles, some people began to protest against the performances of Black Petes: white men who paint their