Teaching High School Students to Use Heuristics While Reading Historical Texts

Jeffery D. Nokes  
Brigham Young University

Janice A. Dole and Douglas J. Hacker  
University of Utah

The purpose of this study was to test the effectiveness of different types of instruction and texts on high schools students' learning of (a) history content and (b) a set of heuristics that historians use to think critically about texts. Participants for the study were 128 male and 118 female students, ages 16 and 17 years, from 2 high schools in the western United States. Eight history classrooms were randomly assigned to 1 of 4 interventions: (a) traditional textbooks and content instruction, (b) traditional textbooks and heuristic instruction, (c) multiple texts and content instruction, or (d) multiple texts and heuristic instruction. The heuristic instruction explicitly taught sourcing, corroboration, and contextualization. Students were administered pretests on their content knowledge and their use of heuristics. After an intervention of 3 weeks, students were readministered the content knowledge and heuristics posttests. A mixed-model analysis of covariance indicated that across all conditions, students who read multiple texts scored higher on history content and used sourcing and corroboration more often than students who read traditional textbook material. Findings highlight the importance of reading multiple texts to deepen content knowledge and facilitate the use of heuristics that historians typically use.

**Keywords:** multiple texts, history, heuristics, strategic readers

**Supplemental materials:** http://dx.doi.org/10.1037/0022-0663.99.X.XXX.supp

The technologies of our information age society are requiring an ever higher degree of literacy from all readers. Typical reading tasks require that readers synthesize information from different texts, critically analyze various sources of information, determine the relevance of different texts, and make judgments between conflicting perspectives. In addition, readers must learn to read in a nonlinear fashion, corroborating information found in one text with that found in different sources and resolving inconsistencies that are often found in multiple documents. One of the places where students have the opportunity to learn these high degrees of literacy is in the study of history. The discipline of history requires historians to do many of the same complex literacy tasks asked of general readers today (Rouet, Favart, Briti, & Perfetti, 1997; Wineburg, 1991b).

Regrettably, in most history classes today, students are not asked to do the kinds of literacy tasks that historians do (Stahl & Shanahan, 2004). Instead, students read from a single history textbook, which they view as the "authority" to be taken at face value. They often read to learn "facts" rather than to critically analyze information (Luke, de Castell, & Luke, 1989; Paxton, 1997; Wineburg, 1991a). The heuristics of the historian—sourcing, corroboration, and contextualization—are rarely taught (Wineburg, 1991b).

The purpose of this instructional study was to evaluate whether high school students could learn to use the heuristics used by historians and also learn history content. We used a quasi-experimental design in which we compared students' learning of history after receiving instruction that varied by the kinds of historical texts that were used (i.e., textbook vs. multiple texts) and by the focus of instruction (i.e., content vs. heuristic). The instructional interventions occurred over a 15-day period and were taught by practicing history teachers in existing classrooms.

**Theoretical Framework**

Traditional comprehension models have focused on how readers understand a single text. For example, Kintsch and van Dijk (1978; see also Kintsch, 2004; van Dijk & Kintsch, 1983) posited the constructive-integrative model of comprehension, which identifies two types of mental representation of a given text. First, readers form a textbase that is a literal representation of the text. Second, readers form a situation model that is an interpretive representation of the text derived from their integration of information from the text with their prior knowledge.

During the 1990s, researchers began to examine the way historians read historical documents, and they determined that the reading of multiple texts required a different comprehension model. Perfetti, Rouet, and Britt (1999) developed what they called a theory of documents representation to describe how good readers process multiple texts, similar to the process used by historians. They used the term documents model to describe readers' mental representation of multiple documents. Perfetti et al. (1999) delineated two components of a documents model. First, there is an intertext model that represents the relationships among and within
the documents themselves and the events described within each. Second, there is a situation model of the total situation described or discussed in all the documents. Together, the intertext model and the situation model make up the documents model. The documents model contains the contents of the texts (similar to a situation model) but also a representation of the sources as well as the connections among the various texts. The documents model is also a mental representation of a text’s relative usefulness and coherence (Wineburg, 1994), including its level of difficulty, trustworthiness, organizational framework, and numerous other textual features.

Perfetti et al. (1999) suggested when expert readers, like historians, read multiple texts, the intertext model and the situation model are in a constant state of interaction and evolution. Historians will judge new information that is read and determine how it is comprehended and interpreted within the larger situation model. The new information drawn from a text informs the situation model and may cause adjustments to be made to it. If a text seems believable and has information that is not currently included in an expert reader’s situation model, the reader may assimilate the new information into that situation model, thus modifying and adding to it. If a text is rhetorically compelling but conflicts with the situation model, an expert reader may accommodate his or her existing situation model to reflect the new information in the text, thus significantly changing the situation model. However, if a text is not compelling yet is drastically different from a reader’s situation model, the reader may view the text as unreliable, and his or her situation model may not change at all. Wineburg (1994) used the analogy of a line-item veto to describe the way an expert reader constructs a situation model. Some ideas from a document are accepted, and others are vetoed as the situation model evolves. Thus, the situation model becomes a cumulative record of the information that has been processed from texts.

**Expert Use of Heuristics and Multiple Documents**

Wineburg (1991a) defined heuristics as “sense-making activities [that help] their user resolve contradictions, see patterns, and make distinctions among different types of evidence” (p. 77). In examining how historians process multiple documents, researchers have determined that historians use several heuristics as they move between their intertext model and their situation model to update their evolving interpretation of events in the texts. Historians approach a reading task with the attitude that *documents are evidence* rather than repositories of facts. Historians use the heuristics of *sourcing*, *corroboration*, and *contextualization* to assist them in developing their situation model of events in the texts.

Using think-aloud protocols, Wineburg (1991a) compared the reading heuristics used by historians with those used by average high school students. Wineburg (1991a) noticed that the historians attempted to construct meaning from multiple documents, but the students focused on remembering facts. Learning the facts proved particularly frustrating for the students because the facts changed from one document to the next. Historians, however, seemed pleased by the contradictions. Differences in the documents brought out the personality of the person who had produced the text, which added a sense of humanness and emotion to the document. Wineburg (1991a) concluded that this recognition of documents as evidence, rather than collections of facts, distinguished experts from novices.

As previously stated, three heuristics also distinguish historians from novice readers of multiple historical texts. One of these is *sourcing*. Sourcing has been operationally defined as the act of looking at a document’s source before reading it and using source information to comprehend and make inferences about the document (Wineburg, 1991a). Historians view primary source documents as interpretations of events through emotional, biased, imperfect witnesses. Secondary source documents introduce additional bias from the historian or journalist who produced them. Thus, to historians, the sources that the reader encounters present distorted images of the past. Documents can only be fully understood and evaluated when the source is kept in mind.

Historians also use *corroboration*. Corroboration is operationally defined as making connections between information found in different texts, with contradictions and similarities being noted (Britt & Aglinaskas, 2002; Wineburg, 1991b). Before accepting an important detail found in one text as plausible, historians check that information against the information found in other texts. Wineburg (1991a) observed that when information in one text contradicted information found in another, historians often turned from the document they were reading back to the text that contained the contradictory information to make direct comparisons.

Finally, historians use *contextualization*. Contextualization is operationally defined as an effort to imagine the particular geographic, political, historical, and cultural context of an event and to comprehend documents within that context (Wineburg, 1991b). Of course, the effectiveness of contextualization is dependent on the historian’s prior knowledge of an event. However, researchers contend that even historians without specific prior knowledge use contextualization through clues in documents (Wineburg, 1996) or by identifying holes in their own knowledge base and recognizing the tentative nature of their interpretations of contextual factors (Britt & Aglinaskas, 2002).

Viewing documents as evidence and use of sourcing, corroboration, and contextualization form the foundation of experts’ processing of multiple texts. These heuristics have been used universally by historians in nearly every study on experts’ processing of multiple texts. Historians’ use of these heuristics is an important part of the historical methodology that historians use, known as *historiography*.

**Novice Use of Heuristics With Multiple Documents**

Research has shown that high school students do not engage in historiography the way historians do. In the study conducted by Wineburg (1991a), 8 academically talented high school students thought aloud as they read the same documents that historians read. The students read the documents in linear fashion, took the information at face value, made more effort to remember the facts than to understand the event, and became frustrated when the documents included contradictions.

Stahl, Hynd, Britton, McNish, and Bosquet (1996) found that high school students were able to develop a basic understanding of a historical event after reading two related documents, but they did not use historiographic heuristics as they read. They failed to notice contradictions between documents and did not use other elements of corroboration or sourcing to help them interpret the
events in the texts. Stahl and his colleagues noted that students lacked the disciplinary knowledge that is fundamental in historiography. They concluded that students needed to be explicitly taught heuristics in high school history classes. Hynd (1999) agreed, pointing out that an opportunity to teach critical thinking skills is lost when history teachers fail to include multiple-text instruction in their classrooms.

In a related study, Britt and Aglinskas (2002) found that neither high school students nor undergraduate students consistently and effectively used souring to critically evaluate multiple documents. For example, both high school and college students referred to material from a novel as if it were factual, a phenomenon also observed by Wineburg (1991a). In addition, merely prompting students to engage in souring did not increase their use of the heuristic. Perfetti, Britt, and Georgi (1995) found that even when college students noticed the sources of multiple documents, they did not use source information as effectively as do historians.

Students' difficulties with multiple texts is not surprising given the fact that they have few opportunities to use them. Most high school history classes rely on a single text, the history textbook. Yet there may be advantages to reading multiple texts rather than the textbook. In two studies, Wiley and Voss (1996, 1999) explored differences in students' writing after reading a textbook passage or multiple texts. Students who read the textbook often repeated the material verbatim, but students who wrote from multiple documents made more transformations of information from the texts. Moreover, the researchers found that writing an argument-based essay after reading multiple texts resulted in the greatest degree of higher order thinking. These results indicate that students are more likely to use heuristics when the reading activity, through the use of multiple texts, implicitly encourages critical thinking and when the writing assignment explicitly asks for individual interpretation.

Recently, researchers have investigated the effects of efforts to teach students to read like historians. Perfetti, MacArthur, and Okolo (2001) found that students as young as 5th-graders became more aware of bias as a factor in the production of historical documents after having experienced several lessons on historical inquiry; however, the researchers admitted in their conclusions that "students did not seem to appreciate the need to take into account the author and his or her motives when reading primary sources" (p. 68). De La Paz (2005) combined explicit instruction in souring and corroboration with explicit instruction on the writing of persuasive essays with 8th grade students. Students who had received this instruction wrote longer, more persuasive, and more accurate essays. However, the results of this study were mixed in terms of students' development of a deeper understanding of historical methodology.

Yet there is some empirical evidence suggesting that high school students can successfully use multiple texts with one or more heuristics to learn historical content. Stahl et al. (1996) and Perfetti et al. (1995) found that although the methods used lacked sophistication, high school students were able to learn historical content from multiple documents. Young and Leinhardt (1998) observed that high school history students' content and disciplinary knowledge deepened as they wrote argumentative, persuasive essays based on multiple primary source documents during the school year. And Britt and Aglinskas (2002) successfully used a computer application called Sourcer's Apprentice to teach students to use souring when reading multiple documents.

In sum, although students rarely use historians' heuristics to analyze multiple documents in history classes, there is evidence to indicate that students at both the high school and college levels seem to be able to learn historical content when reading multiple historical texts. Moreover, using multiple texts, rather than a single textbook, seems to encourage a rudimentary use of at least two heuristics, souring and corroboration. What is necessary, however, is adequate instruction in a scaffolded environment that fosters students' strategic reading so that it resembles that of historians.

Goals of the Present Study

In the present study, we developed four instructional interventions to test the relative effectiveness of heuristic-focused instruction versus content-focused instruction with the use of multiple documents or a traditional history textbook. Eight high school history teachers were randomly assigned to use one of four interventions. We were interested in which intervention led to the greatest gains in content knowledge and whether the heuristics (i.e., souring, corroboration, and contextualization) and the use of documents as evidence could be successfully taught to high school students. Because prior findings suggest that increased learning can be achieved through the use of multiple documents, we expected that classrooms assigned to use multiple documents would show increased learning of historical content over classrooms assigned to use a traditional textbook. Historians' use of heuristics to analyze multiple documents is based on expert knowledge of both content and the discipline. Therefore, whether classrooms assigned to use heuristics could benefit from this type of instruction in the absence of expert knowledge was a question of great interest to us. Additionally, we suspected that using heuristics with a traditional textbook might not represent the best combination of instructional strategies. However, if our results showed potential for this combination, this would be promising news for those instructors who are interested in teaching historiography but may not have access to multiple documents or the time to collect them.

Method

Design

We used a pretest-posttest, quasi-experimental nested design was used in which teacher was treated as levels of a random factor nested within instructional intervention (a fixed factor), and students were treated as a random factor nested within teacher. We determined four instructional interventions by fully crossing two characteristics, type of text (traditional text vs. multiple texts) and type of instruction (content instruction vs. heuristic instruction). Two teachers were randomly assigned to each of the four instructional interventions. Thus, two levels of teacher were nested within each level of instructional intervention.

Participants

Students

Two hundred and forty-six 11th grade students from eight mainstream United States history classes in two high schools in a
suburb of a city in the western United States participated in the study. The selection of the two high schools was based on the similarities of student populations and the willingness of teachers to participate in the study. All students were 16 or 17 years of age. More specific demographics about the students are presented in Table 1, including scores on the American College Testing test and the Stanford Achievement Test, which are standardized tests used to assess students' achievement.

In selecting students for the study, we used several criteria. First, because random assignment of students was not possible, students from existing intact classrooms had to be used. Second, to select a relatively homogeneous group of students across classrooms, we chose mainstream history classes rather than high-ability or low-ability classes. Third, because Jeffrey D. Nokes (J.D.N.), a member of our research team, was an instructor of U.S. history and had considerable expertise in the area of history, the subject selected was U.S. history, an 11th grade subject.

Teachers

The 8 teachers who participated in this study were selected with the help of the social studies specialist at the school district office, who distributed a notice to all high school history teachers in the district, notifying them of a need for teachers to participate in a study. Of the 8 teachers selected, 5 teachers were from one of the selected schools and 3 were from the other. Each of the 8 teachers was rated as "good" by the district social studies specialist. Although several of the teachers taught more than one class of U.S. history, only one class for each of the teachers was selected to participate in the study. There were 3 female teachers and 5 male teachers, and their teaching experience ranged from 2 to 27 years. All 8 of the teachers had bachelor's degrees in history, and 3 had master's degrees. Teachers were randomly assigned to one of the four instructional interventions using a random number generator.

Instructional Interventions

Our goal was to pair a particular type of text with a particular type of instruction. Thus, four instructional interventions were determined by fully crossing two characteristics, type of text (traditional text vs. multiple texts) and type of instruction (content instruction vs. heuristic instruction).

Type of Text

Students used either traditional textbooks or multiple texts that covered the same topics. Efforts were made to keep the length and reading level of all texts equivalent for each reading activity. Table 2 shows the number of words and the Fry readability level of the texts used for each reading lesson. The traditional textbooks consisted of 10 expository texts, one each for each of the 10 reading lessons presented in the study. About half of the expository passages came from 11th grade U.S. history textbooks. For the other half, because textbook accounts inadequately covered some of the topics that were covered by the multiple texts, J.D.N. edited a variety of texts published on the Internet or by historians to make them appear to be expository texts written at a level appropriate for typical 11th-graders (a sample of the texts and the references to all of the texts are available on the Web at http://dx.doi.org/10.1037/0022-0663.99.XXXX.supp).

The multiple texts were selected on the basis of four criteria: First, they had to be at or just below students' reading ability level; second, the various texts had to represent a wide variety of sources, including historical fiction, excerpts from historical speeches, government documents, historical photographs, short excerpts from textbooks, and charts with historical data; third, the multiple texts had to present both agreeing and conflicting points of view; and fourth, the texts had to cover controversial events. The author and source of each text was printed at the bottom of each text.

Because each of the multiple texts focused on a particular topic, the overall collection of texts represented a narrower range of topics than was found in the textbooks. Therefore, once the multiple texts had been identified, the topics contained within them served as the basis on which portions of the traditional texts were selected.

Type of Instruction

The study was embedded in a 15-day history unit on the major events and trends of the United States during the 1920s and 1930s. Within that 15-day unit, half the students received ten 1-hour reading lessons designed to help them learn the historical content in the texts. The other half of the students received ten 1-hour reading lessons that provided instruction designed to help them view documents as evidence and to develop three heuristics: sourcing, corroboration, and contextualization.

The history lesson materials and activities that the teachers used throughout the 15-day unit were developed by J.D.N. These materials included lesson plans, graphic organizers and study guides for students, and keys for teachers. The materials and lessons were identical across all treatment conditions except for the ten 1-hour reading lessons. History activities that were conducted in all classes included discussions, videos, cooperative learning activities, and simulations. Lessons each day were typically divided into two or three activities. These activities were focused on the daily life in the 1920s and 1930s, the Harlem Renaissance, causes and effects of the Great Depression, and foreign affairs of the 1920s and 1930s. Teachers were provided with study guides, graphic organizers, and other materials that were designed to provide

---

Table 1

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Canyon High</th>
<th>Truman High</th>
</tr>
</thead>
<tbody>
<tr>
<td>Student population</td>
<td>2,099</td>
<td>2,242</td>
</tr>
<tr>
<td>White students (%)</td>
<td>95</td>
<td>96</td>
</tr>
<tr>
<td>Hispanic students (%)</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>Students with limited English proficiency (%)</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>Students on free/reduced-price lunch (%)</td>
<td>9</td>
<td>14</td>
</tr>
<tr>
<td>Average daily attendance (%)</td>
<td>98</td>
<td>97</td>
</tr>
<tr>
<td>ACT composite test average</td>
<td>22</td>
<td>21</td>
</tr>
<tr>
<td>SAT reading test average</td>
<td>60</td>
<td>55</td>
</tr>
<tr>
<td>SAT complete battery average</td>
<td>65</td>
<td>58</td>
</tr>
</tbody>
</table>

Note. ACT = American College Testing test; SAT = Stanford Achievement Test.
Table 2
Comparison of the Traditional Texts and Multiple Texts Used in Each Reading Lesson

<table>
<thead>
<tr>
<th>Lesson</th>
<th>Traditional texts</th>
<th>Multiple texts</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Words</td>
<td>Readability*</td>
</tr>
<tr>
<td>1</td>
<td>1,303</td>
<td>8</td>
</tr>
<tr>
<td>2</td>
<td>1,301</td>
<td>11</td>
</tr>
<tr>
<td>3</td>
<td>626</td>
<td>9</td>
</tr>
<tr>
<td>4</td>
<td>1,439</td>
<td>11</td>
</tr>
<tr>
<td>5</td>
<td>2,216</td>
<td>10</td>
</tr>
<tr>
<td>6</td>
<td>1,671</td>
<td>7</td>
</tr>
<tr>
<td>7</td>
<td>1,481</td>
<td>9</td>
</tr>
<tr>
<td>8</td>
<td>859</td>
<td>11</td>
</tr>
<tr>
<td>9</td>
<td>1,107</td>
<td>11</td>
</tr>
<tr>
<td>10</td>
<td>1,592</td>
<td>11</td>
</tr>
<tr>
<td>Average</td>
<td>1,296</td>
<td>9.8</td>
</tr>
</tbody>
</table>

Note. Dashes indicate that the texts consisted of photographs.
* Fry readability level (indicating grade level).

The following reading lesson illustrates the way that lessons were conducted in the first intervention (now this same reading lesson was structured for each of the other three interventions is described below): On the 6th day of the unit, after completing the discussion on the Harlem Renaissance, the teachers in the first intervention passed out a textbook account titled “The Problem of Black Americans.” They passed out a study guide that asked students to answer two literal questions, two inferential questions, and two opinion questions based on the text. In addition, the study guide asked students to write two literal questions, two inferential questions, and two opinion questions based on the text. After writing the questions, students conducted small group discussions related to their questions and answers. The purpose of this activity and all of the other reading lessons used in this first intervention was to help students learn the content of a textbook account that they had read.

**Intervention 2 (multiple texts/content focused)**. Students in this intervention received 10 reading lessons that used multiple texts, with instruction focused on historical content. Because no heuristics were explicitly taught in this intervention, Intervention 2 allowed us to measure the amount of spontaneous use of heuristics when students were provided with multiple texts.

During the 6th day of the unit, after completing the discussion on the Harlem Renaissance, students in the second intervention were given a packet of poems written during the Harlem Renaissance. Students were given a study guide that asked them to answer two literal questions, two inferential questions, and two opinion questions based on some of the poems. The study guide also asked them to write two literal questions, two inferential questions, and two opinion questions based on the poems. After asking the questions, the student were assigned to conduct small group discussions related to each other’s questions. The purpose of this activity and the other reading lessons in this second intervention was to help students learn the content of the multiple texts they had read.

**Intervention 3 (traditional textbook/heuristics focused)**. Students in this intervention received 10 reading lessons that used a traditional text, with instruction on the use of the heuristics. Intervention 3 was used to measure the effectiveness of instruction in heuristics when traditional texts are used. The texts used for each reading lesson were the same expository texts used in Intervention 1. Instruction in this condition typically started with teachers providing direct instruction on the heuristics and modeling the use of the heuristics, followed by opportunities for students to practice the heuristics.

In this particular intervention, we were concerned about the apparent incongruence between the text type and instruction. Was it possible to teach heuristics, such as corroborating and sourcing, using a single textbook account? Typically these heuristics are used with multiple texts, not single texts. But we were interested in whether students would become more strategic in their reading of textbooks without being exposed to multiple documents. In designing lessons, we tried to overcome the apparent incongruence between the texts and the instruction. For example, during one of the lessons on corroborations, the students compared one passage with another part of the same text, looking for internal consistency. In another lesson on corroborations, students compared two textbook accounts. Students also were taught the heuristic of contextualization, and they practiced inferring the social and political context of the time periods they read about in the textbook accounts. Other reading lessons taught students sourcing, by considering the source and purpose of the textbook, and how to use the textbook as evidence to bolster an argument.
On the 6th day of the unit, after completing the discussion on the Harlem Renaissance, students in the third intervention were given a textbook passage titled "The Problems of Black Americans." They were also given a study guide that reminded them about the heuristic of contextualization, which had been introduced to them in an earlier lesson. The teacher reminded them about the heuristic of contextualization by discussing what contextualization is, how it is done, and why it is important. The teacher then modeled for them the use of contextualization as he or she read aloud the first part of the text. The teacher discussed how an individual could get a sense of the values and attitudes of the time from the text. The students then worked on the study guide in small groups. The study guide prompted them to infer from the text the important historical and physical settings of the Harlem Renaissance. Students were asked to infer differences between the values of White and Black America from the text. To conclude, the students were asked to reflect on the heuristic of contextualization and describe why it is important. This reading lesson, along with each of the reading lessons in this intervention, began with a discussion of the heuristic, followed by teacher modeling of the heuristic, and ended with students practicing the heuristic in groups and independently, using an expository text related to the historical theme of the day.

Intervention 4 (multiple texts/heuristics focused). Students in this intervention received 10 reading lessons that used multiple texts, with instruction on the use of the heuristics. The purpose of this condition was to measure students' use of expert heuristics when they are given explicit instruction in the heuristics and opportunities to practice with multiple texts from a variety of genres. The multiple texts used in the fourth intervention were the same texts used in Intervention 2, and the instructional activities were like those used in Intervention 3.

Intervention 4 can be illustrated by examining the 6th day of the unit. After completing the discussion on the Harlem Renaissance, students in this intervention were given the same poems given to students in Intervention 2. The students were also given a study guide. The teacher reminded students about the heuristic of contextualization, which had been introduced to them in an earlier lesson. The teacher discussed with them how and when to use contextualization and why it was important. The teacher then read one of the poems and modeled contextualization by thinking aloud about the values and attitudes that were implied in the poem. Students then worked in small groups analyzing several other poems. For each poem, they were asked to write a general reaction, record what they could infer about the social and political context that surrounded the poem’s creation, and summarize the message the poem was trying to convey. After analyzing five of the poems, they were asked to respond to the following question: "How did the historical context of the Harlem Renaissance influence the writers of these poems and the things they wrote about?" Each of the 10 reading lessons in Intervention 4 followed this same format of having teachers discuss heuristics with students, model the use of one or more of the historians’ heuristics, and give students an opportunity to practice using the heuristic with multiple texts.

Fidelity of Implementation

To ensure that teachers followed the detailed lesson plans, scripts, and activities constructed for the study, J.D.N. observed 84% of classroom sessions and had a formal meeting with each teacher on a weekly basis. During these meetings, J.D.N. provided instruction, answered questions, and received feedback about the implementation of the lessons. These observations and meetings ensured that teachers were faithfully following the lesson plans and scripted materials each day of the 15-day intervention.

Measures

Three dependent measures were used in this study were (a) an observation instrument to measure text use and heuristic instruction during a 3-week baseline period, (b) a measure of students’ learning of the historical content of the 1920s and 1930s unit, and (c) a measure of students’ use of multiple texts and the heuristics.

Observation Instrument

To answer the question of what types of texts and instruction were already in use in the intervention classrooms, J.D.N. and a person independent of the study conducted a 3-week observation of teachers and students in all the intervention classrooms prior to the beginning of the study. J.D.N. designed and piloted the instrument that recorded, at 15-min intervals, the specific types of texts that had been used for at least 5 min and whether multiple texts had been used simultaneously or no text had been used. A second part of the instrument allowed the observer to record, at 15-min intervals, the type of instruction that was taking place and the types of activities in which the class was engaged. Two observers used the instrument while attending 21% of the observed class sessions, and interrater agreement for the two was calculated at 85.9% (the observation instruments are included on the Web at http://dx.doi.org/10.1037/0022-0663.99.X.XXXX supp).

Content Learning

J.D.N. prepared a 40-question, five-option, multiple-choice test to measure students’ learning of the content that was covered during the intervention period. Questions were collected from published past National Assessment of Educational Progress (NAEP) tests and Advanced Placement (AP) tests as well as from tests available through textbook publishers. Additional questions were written by J.D.N. to assess elements of the unit for which appropriate questions could not be found.

In keeping with the aims of both the NAEP and AP tests, many of the questions required students to think at an analytic or inferential level. For example, the following two problems were included on the content test:

Which of the following groups would have been LEAST likely to agree with the nickname “Roaring 20s” to describe the decade of the 1920s?

a. city dwellers,
b. flappers,
c. bootleggers,
d. industrialists, and
e. farmers.
The New Deal changed United States political thinking because it advanced the idea that:

a. businesses should have the right to operate without government interference,

b. government officials should inform the public of major decisions,

c. the President should have a leadership role in determining foreign policy,

d. the federal government should become more involved in the social and economic life of the people, or

e. all of the above.

Interitem reliability of the test was established by administering the test to 12th grade students (N = 49) who were not part of the study but who had taken U.S. history previously. Reliability using Cronbach’s alpha was calculated at .89. Content validity was established by ensuring that test items were representative of those on the standardized NAEP and AP tests and also reflected content covered in the lessons. In addition, construct validity was calculated by correlating the scores of the 246 students who were part of this study with their score on the social studies component of the SAT. The Pearson coefficient was calculated at .49 (p < .001). According to Anastasi (1988), a moderately high correlation for construct validity is desirable (the content test is available on the Web at http://dx.doi.org/10.1037/0022-0663.99.X.XXX.supp).

**Use of Heuristics**

To measure students’ pretest and posttest use of the heuristics during document analysis, J.D.N. used a method similar to that used by Wineburg (1991a). This method consisted of three stages:

First, students were given multiple documents and a picture related to a single event and told to critique the picture on the basis of the information in the documents. Second, students were asked to take on the role of a historian and write a 200-word essay explaining whether they believed the picture accurately portrayed the event discussed in the documents. Third, students were asked four open-ended questions that were intended to reveal their use of heuristics. These questions were:

1. Which two documents would you rank as the most reliable? Why would you rank them as most reliable?

2. Which two documents would you rank as the least reliable? Why would you rank them as least reliable?

3. Which two documents would you rank as the most helpful in writing your essay? Why would you rank them as most helpful?

4. Which two documents would you rank as the least helpful in writing your essay? Why would you rank them as least helpful?

These three stages were administered twice to students, once as a pretest and once as a posttest. The pretest and posttest used two different topics, the Battle of Lexington and the Pullman Strike, neither of which was studied during the 1920s and 1930s unit. The two topics were counterbalanced so that half of the students in each class received the Battle of Lexington as the protest and the other topic as the posttest. For the other half of the students, the order of the topics was reversed.

Students’ use of sourcing, corroboration, and contextualization and their use of documents as evidence on the pretest and posttest were measured using a coding scheme that J.D.N. prepared. On the basis of research by Britt and Aglinskis (2002) and Wineburg (1991a, 1994), we constructed a coding sheet as well as detailed coding instructions that operationally defined each heuristic, listed descriptors of what should and should not be considered an instance of heuristic use, and gave examples of heuristic use (the two forms of the heuristics test, the coding sheet, and the complete set of coding instructions are available on the Web at http://dx.doi.org/10.1037/0022-0663.99.X.XXX.supp).

To illustrate how evaluators scored students’ answers for the presence of heuristic use, the following instructions were given for sourcing:

As individual who uses sourcing looks at the source of a document before reading and keeps the source of the document in mind as he or she reads. The reader’s understanding of the document is influenced by the document’s source. Sourcing only occurs when the consideration of the source helps the individual make sense of the document. If the student analyzes two documents together, give two marks (For example, if the student was to write “Both Document 2 and Document 7 are biased because the authors of both documents wanted to blame the other side for the event.”)

In addition, the evaluators were given eight subheadings listing specific content they should look for and three subheadings listing content that should not be considered sourcing. One of the eight subheadings described using author’s position in sourcing as follows:

Any reference to the occupation, profession, level of training, or other credentials of the author of the document in order to suggest that the document is more or less reliable or in order to understand what the document says, qualifies as sourcing. Examples: “Since Shaw was an officer in the British army, he would have known...”, “The historian who wrote this must have studied a lot to become a historian so...”

The heuristics test was piloted with a group of 11th and 12th grade students (N = 31) who were not part of the study. These students’ essays were used to develop and test the coding instructions and coding sheet. As students’ essays were assessed, the evaluators tallied the number of times the students provided evidence of heuristic use in their writing. Students received a score on each heuristic based on the number of times they used it in their writing. The pilot tests helped the evaluators work through the specific factors that determined when a student would be given credit for heuristic use.

Two evaluators were trained to use the coding sheet and the coding instructions using several sample tests that they evaluated together. Following their training, both evaluators assessed 20% of the pretests and posttests. They were in complete agreement on 78% of the heuristic scores and within one point on 95% of the heuristic scores. Evaluators conducted a blind assessment of the tests, unaware of the intervention they were assessing or whether each test was a pretest or a posttest.
Procedure

Prior to the administration of the instructional intervention, a 3-week observation period took place during which the 8 teachers were observed and the observation instrument was used to record the types of texts they used and the type of instruction they provided. Following the 3-week observation, the 8 history teachers administered the content prettest and the heuristic essay prettest to the students in their classes. Teachers then conducted the intervention using the researcher-prepared history activities and intervention reading lessons during the 15-day unit on the 1920s and 1930s. Following this, the teachers administered the content posttest to measure students' learning of the historical information of the unit. Finally, students took the heuristic essay posttest to assess their ability to use the heuristics that historians typically use.

Results

Observations

The 3-week observations indicated that none of the 8 teachers provided opportunities for students to read multiple texts on a regular basis, nor did they teach heuristics. In no instance did teachers take time to carefully analyze multiple sources or discuss with students heuristics for considering multiple sources or how heuristics could be used to improve comprehension. Instead, teachers typically engaged in traditional activities, spending more time lecturing than engaging students in any other type of activity. They relied primarily on the textbook and study guides as the texts of choice in their classes. There were differences among the teachers in how they used textbooks; however, these differences were not directly related to the conditions manipulated in the study.

History Content Test

The primary analysis used for this study followed a statistical procedure that has been widely accepted in quasi-experimental classroom studies in which a nested analysis of covariance (ANCOVA) is used (Hopkins, 1982; Kromrey & Dickinson, 1996). First, using teacher as the unit of analysis, we conducted a fully specified mixed-model ANCOVA with student (a random factor) nested within teacher and with teacher (a random factor) nested within intervention (a fixed factor). This analysis provided a statistical test of the null hypothesis for the teacher-within-intervention effect. If this test was not statistically significant for teacher, the individual teacher could be eliminated as a "nuisance" factor, and the unit of analysis could be turned from teacher to individual student, thereby providing greater statistical power to detect intervention effects. A liberal alpha level is typically recommended to test the teacher-within-intervention effect. Recommended alpha levels range from .10 (Winer, 1971) to .20 (Hopkins, 1982) to .25 (Kirk, 1982) to .30 (Critt-Christoph & Minz, 1991). For the present study, we compromised and selected an alpha of .20. Thus, when we found no teacher-within-intervention effect, teachers within an intervention were combined, and individual students became the unit of analysis. We conducted an ANCOVA in which the pretest scores were used as the covariate and the posttest scores were used as the dependent variable. We compared differences by conducting post hoc multiple comparison tests on unadjusted means, and we calculated effect sizes using partial eta-squared.

Table 3 shows the unweighted pretest and posttest means and standard deviations for students' content test scores for each classroom and intervention. A mixed-model ANCOVA, which used teacher as the unit of analysis, showed no significant teacher-within-intervention effect (Wald $z = .66, p = .51$). An ANCOVA, which used students' scores as the unit of analysis, showed a significant intervention effect, $F(3, 213) = 21.93, p < .001$, partial $\eta^2 = .24$. A follow-up multiple comparison test using Tukey's honestly significant difference (HSD) showed that students who used multiple texts to study content scored significantly higher on the history content posttest than did all other groups ($p < .01$). In addition, students who used multiple texts to study heuristics scored significantly higher than students who used traditional texts to study heuristics ($p < .02$). Figure 1 illustrates these differences.

Thus, higher performance on the history content test appears to have been the result of using multiple texts versus traditional texts.

Heuristics Essay Test

The heuristics essay test was used to determine students' ability to use the heuristics of sourcing, corroboration, and contextualization. In addition, students' were evaluated on their ability to cite the documents as evidence in their writing. Students received a score for each of these heuristics on the pretest and posttest based on the number of times there was evidence of heuristics use in their writing.

Sourcing

Table 4 shows the unweighted pretest and posttest means and standard deviations for sourcing for each intervention. A mixed-model ANCOVA, which used teacher as the unit of analysis, showed no significant teacher-within-treatment effect (Wald $z = .84, p = .40$). An ANCOVA, which used students' scores as the unit of analysis, showed a significant intervention effect, $F(3, 206) = 16.35, p < .001$, partial $\eta^2 = .19$. A follow-up multiple comparison test using Tukey's HSD showed that students who used multiple texts to study heuristics scored significantly higher on the sourcing posttest than did all other groups ($p < .001$; see Figure 2). There were no other significant differences. Moreover, more students used sourcing than any of the other heuristics, with 155 (76%) using it 1–15 times, averaging about 3 times by each student.

Corroboration

Table 4 shows the unweighted pretest and posttest means and standard deviations for corroboration for each intervention. A mixed-model ANCOVA, which used teacher as the unit of analysis, showed no significant teacher-within-treatment effect (Wald $z = 1.19, p = .23$). An ANCOVA, which used students' scores as the unit of analysis, showed a significant intervention effect, $F(3, 205) = 10.02, p < .001$, partial $\eta^2 = .13$. A multiple comparison test using Tukey's HSD showed that students who used multiple texts to study heuristics scored significantly higher on the corroboration posttest than did the two groups that worked with textbooks ($p < .01$). In addition, the group that used multiple texts to study content scored significantly higher on the corroboration posttest than the group that used the textbook to study content (see...
Table 3

Unweighted Mean Pretest and Posttest Scores (With Standard Deviations in Parentheses) on Content Test by Intervention and Classroom

<table>
<thead>
<tr>
<th>Intervention and classroom</th>
<th>n</th>
<th>Pretest</th>
<th>Posttest</th>
</tr>
</thead>
<tbody>
<tr>
<td>Traditional textbook/content focused</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Classroom 4</td>
<td>33</td>
<td>16.17 (5.07)</td>
<td>25.67 (6.44)</td>
</tr>
<tr>
<td>Classroom 8</td>
<td>31</td>
<td>17.58 (5.01)</td>
<td>25.08 (4.43)</td>
</tr>
<tr>
<td>Combined</td>
<td>64</td>
<td>16.85 (5.05)</td>
<td>25.38 (5.53)</td>
</tr>
<tr>
<td>Multiple texts/content focused</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Classroom 2</td>
<td>31</td>
<td>14.96 (5.33)</td>
<td>30.55 (3.91)</td>
</tr>
<tr>
<td>Classroom 3</td>
<td>31</td>
<td>13.40 (6.40)</td>
<td>30.43 (5.48)</td>
</tr>
<tr>
<td>Combined</td>
<td>62</td>
<td>14.19 (5.89)</td>
<td>30.49 (4.72)</td>
</tr>
<tr>
<td>Traditional textbook/heuristics focused</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Classroom 6</td>
<td>29</td>
<td>15.16 (4.50)</td>
<td>24.93 (6.51)</td>
</tr>
<tr>
<td>Classroom 7</td>
<td>30</td>
<td>13.13 (5.14)</td>
<td>20.63 (6.70)</td>
</tr>
<tr>
<td>Combined</td>
<td>59</td>
<td>14.13 (4.90)</td>
<td>22.74 (6.90)</td>
</tr>
<tr>
<td>Multiple texts/heuristics focused</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Classroom 1</td>
<td>31</td>
<td>15.21 (5.94)</td>
<td>27.23 (6.50)</td>
</tr>
<tr>
<td>Classroom 5</td>
<td>30</td>
<td>14.43 (4.99)</td>
<td>25.62 (5.73)</td>
</tr>
<tr>
<td>Combined</td>
<td>61</td>
<td>14.82 (5.46)</td>
<td>26.44 (6.14)</td>
</tr>
</tbody>
</table>

Figure 3. Corroboration was used much less often than sourcing in all interventions, and it was quite rare in those groups that had worked with the textbook rather than multiple documents.

Contextualization

Because only 15 students (7%) used contextualization, no analyses were possible. This heuristic was used rarely by students in all interventions, both before and after the interventions.

Using Documents as Evidence

Table 4 shows the unweighted pretest and posttest means and standard deviations for using documents as evidence for each intervention. A mixed-model ANCOVA, which used teacher as the unit of analysis, showed no significant teacher-within-treatment effect (Wald $z = .97$, $p = .33$). An ANCOVA, which used students' scores as the unit of analysis, was not statistically significant, $F(3, 205) = 1.51$, $p = .21$. No other analyses were conducted with this variable. In sum, many students cited documents in their writing. However, there was little difference between students in different interventions, and there was little difference between pretest and posttest scores.

Discussion

The purpose of this study was to compare students' learning of historical content and use of historians' heuristics using four...
Table 4.  
Unweighted Mean Pretest and Posttest Scores (With Standard Deviations in Parentheses) on Sourcing, Corroboration, and Using Documents by Intervention

<table>
<thead>
<tr>
<th>Intervention</th>
<th>n</th>
<th>Pretest</th>
<th>Posttest</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Sourcing</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Traditional textbook/content focused</td>
<td>64</td>
<td>2.21 (3.11)</td>
<td>1.94 (2.19)</td>
</tr>
<tr>
<td>Multiple texts/content focused</td>
<td>62</td>
<td>2.39 (2.46)</td>
<td>2.52 (1.95)</td>
</tr>
<tr>
<td>Traditional textbook/heuristics focused</td>
<td>59</td>
<td>1.33 (1.90)</td>
<td>1.76 (1.81)</td>
</tr>
<tr>
<td>Multiple texts/heuristics focused</td>
<td>61</td>
<td>1.93 (2.00)</td>
<td>4.39 (2.83)</td>
</tr>
<tr>
<td>Total</td>
<td>246</td>
<td>1.97 (2.44)</td>
<td>2.66 (2.45)</td>
</tr>
<tr>
<td><strong>Corroboration</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Traditional textbook/content focused</td>
<td>64</td>
<td>0.40 (0.77)</td>
<td>0.22 (0.53)</td>
</tr>
<tr>
<td>Multiple texts/content focused</td>
<td>62</td>
<td>1.44 (1.45)</td>
<td>0.76 (1.06)</td>
</tr>
<tr>
<td>Traditional textbook/heuristics focused</td>
<td>59</td>
<td>0.52 (0.70)</td>
<td>0.36 (0.85)</td>
</tr>
<tr>
<td>Multiple texts/heuristics focused</td>
<td>61</td>
<td>0.44 (0.71)</td>
<td>1.04 (1.18)</td>
</tr>
<tr>
<td>Total</td>
<td>246</td>
<td>0.58 (1.01)</td>
<td>0.59 (0.99)</td>
</tr>
<tr>
<td><strong>Using documents</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Traditional textbook/content focused</td>
<td>64</td>
<td>1.47 (1.36)</td>
<td>1.24 (1.23)</td>
</tr>
<tr>
<td>Multiple texts/content focused</td>
<td>62</td>
<td>1.92 (1.48)</td>
<td>1.74 (1.45)</td>
</tr>
<tr>
<td>Traditional textbook/heuristics focused</td>
<td>59</td>
<td>1.46 (1.49)</td>
<td>1.16 (1.34)</td>
</tr>
<tr>
<td>Multiple texts/heuristics focused</td>
<td>61</td>
<td>1.71 (1.45)</td>
<td>1.76 (1.65)</td>
</tr>
<tr>
<td>Total</td>
<td>246</td>
<td>1.64 (1.45)</td>
<td>1.48 (1.44)</td>
</tr>
</tbody>
</table>

different instructional interventions. In these interventions, we manipulated whether students read multiple texts or a traditional textbook and whether the instructional focus was on content or historiography. When tested on historical content, students who read multiple texts with a focus on heuristics outperformed those who read traditional texts with a focus on historical content or heuristics. Thus, using multiple texts as an alternative to a traditional history textbook resulted in superior learning of historical content, regardless of whether the focus of instruction was on content or heuristics. These results are mirrored in studies by Stahl et al.

![Interventions](image)

Figure 2. Mean scores for using sourcing heuristic on pretest and posttest by intervention.
(1996) and Perfetti et al. (1995), who showed that the use of multiple texts led to gains in students’ content knowledge.

These findings suggest a number of implications for the teaching of history in high school classrooms. First, the value of multiple texts appears to be the most important finding of the study. Across all conditions, students who read multiple texts learned more than students who read single texts. Unfortunately, the single history textbook continues to be the most commonly used text in history classes (Paxton, 1999). Teachers may believe that the only way for students to master the required historical content is to present it through a concisely written single source (Snow, 2002). Coupled with this, there may be a belief that using multiple sources actually detracts from learning the required content. However, the present study suggests that a collection of thoughtfully selected sources can actually enhance the learning of historical content. In addition, although student affective states during the interventions used in this study were not measured, other studies have shown that using multiple documents in the classroom positively impacts student engagement and motivation (Guthrie & Cox, 1997; Guthrie et al., 1996, 2000).

Second, the teaching of heuristics in high school history classes may have value in helping students learn the heuristics that historians use as they think about and write history. Teachers in the study were able to successfully teach the heuristics of sourcing and corroboration so that students could use these in their own writing. Further, students used documents as evidence many times throughout their written essays. The use of these heuristics could be a first step in teaching students how to gain a deeper understanding of history and to think about and write about history more in the way that historians do.

However, teachers were not successful in teaching their students to use contextualization, even though this heuristic is a key to comprehending historical texts and a heuristic that historians use consistently. It may be that high school students do not have the depth of background knowledge in history to use this heuristic, that this heuristic takes longer to learn, or that it was not taught with enough potency. It may be that the heuristic is too inherently difficult to teach to high school students or to students at any age level. Regardless, research has shown that historians make ready use of this heuristic and find it essential to generating interpretations of historical events (Wineburg, 1991a, 1998); therefore, future intervention and classroom investigations are warranted.

Findings from this study suggest that history teachers can include heuristics instruction with multiple texts in their classrooms without compromising students’ learning of historical content. In fact, students who received heuristics instruction with multiple texts scored significantly higher on the content posttest than their peers who were part of the more traditional intervention, using textbooks to study content. As a result, this study provides empirical support for the increasing call for history teachers to include historiography instruction along with traditional instruction in their classrooms (Stahl & Shanahan, 2004). However, historical content was learned significantly better by students who read multiple texts and were taught the historical content than by students who read multiple texts and learned the heuristics. So, teachers need to be careful in balancing the teaching of historical content and heuristics. In fact, a pressing issue for all teachers is the thoughtful balance between teaching the learning process and teaching important content.

**Limitations**

There are several limitations of this study. First, it was designed to examine the effects of the instructional interventions, not the teachers. Thus, as in many quasi-experimental studies, there was a need in this study to reduce teacher effects. To do this, we scripted all lessons and asked teachers to follow the scripts. As a result, teachers were given little flexibility in the implementation of lesson plans. They were allowed to answer students’ questions and to elaborate when they judged it was necessary to help students learn principles; but beyond this, there was little room for
teachers to adjust lesson plans to fit their styles. Thus, one of the trade-offs in this study was that we were able to establish treatment effects that outweighed teacher effects by not allowing teachers to flexibly implement lesson ideas.

An additional limitation of this study was our analysis of students' use of heuristics rather than an analysis of students' deep historical thinking. Ultimately, research in this field should focus on helping students engage in deep historical analysis. Certainly the use of heuristics is a means to this end but not an end in itself. However, students' use of sourcing, corroboration, and contextualization serves as an important bridge between the unsophisticated reading in which students typically engage and the sophisticated, deep reading in which historians engage. Research suggests that the use of the heuristics of sourcing, corroboration, and contextualization is a prerequisite for sophisticated historical analysis. This study provides important data about teaching students to use these heuristics. It does not provide data on the relationship between the use of the heuristics and students' engagement in deep historical analysis, which is the ultimate goal of the heuristics instruction.

Conclusion

Certainly, living in a democratic society during the information age demands that citizens be critical readers of the barrage of information that inundates them on a daily basis. Rarely, if ever, is there a single source—such as the traditional history textbook—that concisely encapsulates the complex issues facing our society. Rather, there are a variety of sources to be read, evaluated, synthesized, and interpreted. Therefore, exposing students to a variety of sources and educating them in the use of heuristics, which they can use to generate their own interpretations, seems more like a necessity than an add-on that is offered only to a few students in advanced placement courses. The results of this study hopefully will motivate history teachers to carefully reconsider their approaches to classroom practices. We all need the heuristics of a historian. What better place to learn them than in a history classroom?

References


Received June 15, 2006
Revision received April 8, 2007
Accepted April 11, 2007