Developing Historical Reading and Writing With Adolescent Readers: Effects on Student Learning

Susan De La Paz
University of Maryland

Mark Felton
San José State University

Chauncey Monte-Sano
University of Michigan

Robert Croninger, Cara Jackson, Jeehye Shim Deogracias, and Benjamin Polk Hoffman
University of Maryland

Abstract: In this study, the effects of a disciplinary reading and writing curriculum intervention with professional development are shared. We share our instructional approach and provide writing outcomes for struggling adolescent readers who read at or below basic proficiency levels, as well as writing outcomes for proficient and advanced readers. Findings indicate significant and meaningful growth of about 0.5 of 1 standard deviation in students’ abilities to write historical arguments and in the length of their essays for all participants, including struggling readers. Our study also considers teacher implementation of the curriculum intervention. We found that teachers who were most faithful to the underlying constructs of our curriculum intervention also made successful adaptations of the lesson materials.

Cara Jackson is now at Urban Teacher Center in Baltimore, MD. Jeehye Shim Deogracias is now at Alexandria City Public Schools in Alexandria, VA. Benjamin Polk Hoffman is now at the University of Houston in Houston, TX.

Correspondence should be sent to Susan De La Paz, Department of Counseling Psychology, Higher Education and Special Education, College of Education, University of Maryland, 3216 Benjamin Building, College Park, MD 20742. Email: sdelapaz@umd.edu
Developing Historical Reading and Writing

Keywords: disciplinary literacy, historical arguments, adolescent readers, cognitive apprenticeships, student learning outcomes

Educators and researchers have long called for teaching with primary sources and inquiry methods in history classes (Bain, 2006; Wineburg, 2001), an approach that relies on historical ways of reading, thinking, and writing. Although there has been some attention to these ideas in high school and college history classrooms (Monte-Sano, 2008; Nokes, Dole, & Hacker, 2007; Reisman, 2012b; Stahl, Hynd, Britton, McNish, & Bosquet, 1996; Wiley & Voss, 1999; Young & Leinhardt, 1998), we have found minimal attention to middle school students (for an exception, see De La Paz [2005]). In addition, educators have drawn attention to the needs of struggling readers and writers (Biancarosa & Snow, 2006; Graham & Perin, 2007; McCombs, Kirby, Barney, Darilek, & Magee, 2005). Yet, typical classrooms are academically diverse, requiring teachers to satisfy the needs of both struggling and successful students at the same time. We constructed an intervention to address many of these demands while embracing current thinking about teaching history and writing. We placed historical argumentation at the center of our efforts, believing that it integrates many of the pressing goals for students, such as historical thinking, disciplinary ways of reading and writing, and argumentation.

History is an interpretive discipline grounded in evidence-based argument. The ability to craft historical arguments has become a central feature of national standards like the Common Core State Standards and C3 Framework (Council of Chief State School Officers [CCSSO], 2010; National Council for the Social Studies [NCSS], 2013). To make historical arguments, students must attend to conventional features of argumentation, such as constructing a claim, integrating relevant evidence, and explaining how that evidence supports their claim (Toulmin, 1958). Ideally, students also recognize and address counterclaims or counterevidence. In addition to being a generic concept, applicable to many subject areas, argumentation is crucial to history in its orientation to interpretation and evidence. In contrast to what many students believe, history is not a given set of fixed information, rather it is a deliberative process of constructing arguments grounded in the available evidence (Holt, 1990; Mink, 1987). To engage with the discipline requires historical ways of reading, thinking, and writing. We use the concept of argumentation to capture, in part, the evidence-based interpretive nature of history, recognizing that it does not address all aspects of historical accounts (Lee, 2005).

Yet history requires more than argumentation alone. Our central learning goal also embraces the disciplinary thinking and understanding rooted in historical concepts and practices (Lee, 2005; Seixas, 2006; VanSledright & Limón, 2005). As historians study the past, they apply concepts, such as time, change, context, empathy, cause, and evidence to their analyses. They engage in such procedures as researching, critiquing sources of evidence, or constructing...
interpretations. Historical concepts and practices shape arguments that have disciplinary integrity.

And since history rests on analysis of sources and sharing conclusions in writing, particular ways of reading and writing are embedded in historical study. But these are not necessarily general ways of reading and writing that can be applied to any discipline as defined in much of the early literature on content area literacy literature (Biancarosa & Snow, 2006). And we seek to extend beyond efforts that focus on using discipline-specific texts and writing activities to enhance content area learning (e.g., Daniels, Zemelman, & Steineke, 2007). Here, we look to more recent conceptions of disciplinary literacy (Moje, 2008; Shanahan & Shanahan, 2008) that highlight the importance of moving beyond basic and intermediate literacy instruction to focus on discipline-specific ways of reading and writing and the kinds of disciplinary thinking embedded in them. When it comes to reading, we consider Wineburg’s (1991) identification of sourcing, corroboration, and contextualization, as historical ways of reading and analysis, and De La Paz and Felton’s (2010) and Monte-Sano’s (2010) ideas about the disciplinary use of evidence in students’ historical essays. In other words, we focused on ways of reading, thinking, and writing that are foundational to historical analysis rather than more general literacy strategies that might apply to many content areas. We strove to teach adolescents to question historical sources and, when writing, to develop sophisticated claims by explaining the relevance of a quote, then discussing it, illustrating it with examples, or using it to counter an opposing claim. Finally, we hoped students would also learn to establish the evidentiary strength of their claims. These goals were based in part on past findings that teachers can help students to develop evidence-based writing by assigning argument and analytical writing in combination with reading multiple texts (Monte-Sano, 2010; Wiley & Voss, 1999; Young & Leinhardt, 1998).

We surveyed this literature and built on past research experience to target historical argumentation, in general, and perspective, contextualization, and substantiation, in particular, as expressed in student essays. Argumentation refers to the interpretive nature of history and substantiation to the evidence-based nature of those arguments. Perspective highlights attention to people’s views in the past, including authors of texts and their credibility. Contextualization requires students to ground their arguments in the time and place under consideration, avoid anachronism, and identify relationships between events. Conveying these historical arguments in writing reflect the way in which historians share their conclusions as well as important argumentative writing goals for middle school students. Of course, these are just some of the disciplinary literacy practices that represent history, but they do reflect the key emphasis on informational texts and argumentation in the Common Core State Standards as well as inquiry and historical thinking in the C3 Framework (CCSSO, 2010; NCSS, 2013).
We defined these goals so they were precise enough to be measured, narrow enough to be addressed deliberately in an 18-day curriculum, realistic for struggling readers, and appropriate for students and teachers who had largely not been exposed to an inquiry-oriented, literacy-rich history classroom. We found that such an approach allowed both struggling and successful middle school students to gain a foothold in historical thinking and evidence-based argumentative writing, addressing several concerns reformers have expressed over the years. These goals have historically been more readily attainable for older learners, and thus we pause to elaborate further on the needs of novice and struggling students and to situate our approach to teaching and learning.

**Challenges for Novice and Struggling Readers and Writers**

To write evidence-based arguments from primary source documents, students must engage in a host of interrelated activities driven by the purpose of the writing task. Without a big-picture representation of the process, they often get lost in one of the many subroutines that they enact. Students must interpret the writing prompt, read documents for information and evidence as it relates to the prompt, and write with a rhetorical plan that organizes and reconciles the evidence that they have collected from the documents (Young & Leinhardt, 1998). It is a complex process that places high demands on students with learning disabilities (LDs) and other learners who struggle academically, made all the more difficult for students who bring less content knowledge to subject area learning (Mastropieri, Scruggs, & Graetz, 2003).

To begin, struggling readers and writers often have difficulty discerning the underlying goals and purposes of writing tasks. Without direct instruction embedded in content, they misrepresent the purpose of writing, or more often, write without elaborated goals to guide the planning and composing process (Graham, 2006), although there is evidence to suggest that when given explicit directives in goal setting, students can write more complete arguments (Ferretti, Lewis & Andrews-Weckerly, 2009; Page-Voth & Graham, 1999). Moreover, a consistent finding is that the degrees to which students know about the topic influences their ability to comprehend and reason with multiple documents (Rouet, Britt, Mason, & Perfetti, 1996; Rouet, Favart, Britt, & Perfetti, 1997; Stahl et al., 1996; Young & Leinhardt, 1998). And when reading historical texts, students often focus on the literal meaning of documents and miss intertextual reading strategies that would support interpretive and analytic work (Afflerbach & VanSledright, 2001; Stahl et al., 1996). Thus, novice and struggling readers are more likely to have trouble making inferences, or recognizing subtext when reading historical artifacts, and less likely to then evaluate documents as evidence. Finally, argumentative writing calls for new text structures that require greater organizational and linguistic processing than
is required from genres emphasized earlier in the school curriculum (Coirier, Andreissen, & Chanquoy, 1999; Crowhurst, 1991). Thus, struggling writers need support in developing both the text structures for argument and the linguistic devices to express them. Given the heavy demands associated with producing argument structures, struggling writers need support with regulating the writing process (Felton & Herko, 2004).

Another cognitive task is that the ability to interpret primary accounts of historical events also requires knowledge of specific historical contexts, based on particular facts regarding the time and place of the events (Van Drie & Van Boxtel, 2008). In other words, when reading, students must interpret evidence in the account in accordance with its context (Halldén, 1997). Adolescents have been observed to have difficulty grasping the nature of historical context (Husbands, 1996; Shemilt, 1983), tending to judge past actors and actions by present standards (VanSledright, 2002).

Much of the research has focused on how students read and cite evidence from multiple source documents (J. C. Greene, 2001; Young & Leinhardt, 1998). Some additional attention has been granted to how students represent and construct historical arguments (Britt, Rouet, Georgi, & Perfetti, 1994; Wiley & Voss, 1999). But more research is needed to understand how students develop proficiency with these skills. Further, with few exceptions (Montesano, 2008), studies of writing development in history have focused on college students (Britt et al., 1994; S. Greene, 1993; Wiley & Voss, 1999) or high school students in advanced placement courses (Young & Leinhardt, 1998). Studies that compare learning outcomes for average or struggling secondary learners are rare (De La Paz & Felton, 2010).

While there is a growing base of research in disciplinary literacy to build on, little is known about instructional interventions designed specifically for struggling readers and writers in middle school history classrooms. Given the complexities in writing historical arguments based on primary source documents, and the varying learning needs of students in heterogeneous classrooms, we believe it is critical to provide students with a supportive learning environment that is both holistic and comprehensive in its approach. Struggling readers need to develop deep understandings of the connections between reading and writing, they need time to practice analyzing and evaluating texts, and they need to learn how to plan and develop arguments, as they coordinate interrelated ways of thinking to accomplish the more global goal of writing historically.

**Successful Models of Literacy Instruction**

The kinds of literacy and historical thinking practices we target are not necessarily obvious to those lacking related experience. Thus, it is unfair to expect students to acquire these skills without clear expectations. One helpful
model for instruction is cognitive apprenticeship (Brown, Collins, & Duguid, 1989), which makes expert thinking and literacy practices visible to novices through teacher modeling. As students gain practice in the new ways of reading, thinking, and writing, less modeling is required, but regular feedback is still needed to support learning so that students are able to use these thinking and literacy practices independently. In fact, a cognitive apprenticeship is both an instructional model that teachers use to organize the learning environment and an approach to learning that helps students to see the processes involved in complex learning activities. Brown and his colleagues (1989) envisioned that cognitive apprenticeships could focus on complex, higher-order thinking; make visible heuristic strategies used by experts; model, coach, and scaffold for novices as they learn to how to use those heuristics; and finally, through an evolving community of practitioners, achieve the habits of critical thinking.

Cognitive apprenticeships have long been found to be a successful literacy model for teaching young and struggling readers and writers to compose generic writing forms, such as narrative, persuasive, and expository essays (e.g., Harris, Graham, & Mason, 2006). In this work, teachers situate writing as a purposeful activity and apply several heuristics in an expert–novice apprenticeship, which have been found effective in recent meta analyses (Graham & Perin, 2007). This approach has shown promising results for students with varying ability levels, including those who progress without difficulty in the curriculum and students who are gifted (Albertson & Billingsley, 2001; Glaser & Brunstein, 2007) and in a limited number of studies that focus on disciplinary literacy (e.g., De La Paz, 2005; Engle & Conant, 2002). Our approach is consistent with Engle and Conant’s (2002) guiding principles for fostering productive disciplinary engagement, as we ask that teachers engage students in problematizing the content, giving students authority to solve problems, holding them accountable to others and disciplinary norms, and providing them resources to engage in the activities.

Here, we employ these principles through a series of historical investigations where disciplinary questions were posed for students to answer using primary source documents, culminating in essays in which they created historical arguments and evidence to draw their own conclusions. Using the gradual release model also created an environment where peers often collaborated in reading, thinking, and writing processes. Thus, there are many similarities between our approach to teaching and earlier cognitive apprenticeship models, such as having teachers situate the purpose for learning within a meaningful and authentic context (exploring controversial historical investigations), having teachers guide the learning process by demonstrating a coordinated approach to the learning task (e.g., by reflecting with students how to consider multiple aspects in evaluating an author’s perspective), and having teachers solicit input from students through collaborative work in a close reading of documents.

Our instructional model differs from prior cognitive apprenticeship in several ways. In particular, we focus the cognitive apprenticeship on learning
disciplinary literacy strategies in history rather than more general strategies, as has been customary. We also place special emphasis on two additional aspects of apprenticeship: we prioritize the need for students to attempt to gain control over their own learning by providing multiple opportunities for independent application of cognitive routines, and we privilege the need for teachers to give students feedback on their learning attempts. Teachers systematically decrease the amount of whole-class instruction over the course of the instructional program, and they respond more to individuals and small groups within the classroom, as they work to meet the varying needs found among diverse groups of students who are learning how to annotate documents and evaluate evidence and to plan mindfully before composing. Our approach to instruction differs in important ways from some forms of strategy instruction that focus more closely on discrete skills. First, we show learners how to engage with the content in self-determined ways and with varying degrees of sophistication. Second, our curriculum asks students to integrate reading and writing as they think about historical content, to use a set of historical thinking heuristics in a coordinated fashion, and privileges learning about history. In sum, we use a cognitive model as the basis for our apprenticeship but consider guiding principles for disciplinary engagement in history classrooms.

**OVERVIEW OF THE STUDY AND RESEARCH QUESTIONS**

The purpose of this study was to gain preliminary data on the effectiveness of a cognitive apprenticeship approach to teaching academically and culturally diverse middle school students to compose historical arguments. We developed an 18-day curriculum intervention for students, evaluating students’ essays written before and after the intervention. Moreover, we compared writing from students whose teachers used our approach to writing from students whose teachers were not involved in the project. Finally, because cognitive apprenticeship requires a dynamic approach to instruction where teachers must gradually release responsibility for strategy use to the students, we looked at teacher fidelity to the intervention. These data provided both a validity check on the intervention and a means of assessing the direct impact of the intervention on student outcomes. This study shares data from the first year of a three-year project that involved close examination of how teachers and students responded to the curriculum, and we used the results to refine both the intervention and the professional development (PD). We asked several research questions to explore the relationships in teaching and learning.

1. What are the effects of a historical thinking curriculum intervention with teacher PD on the disciplinary writing skills of culturally and academically diverse students?
Developing Historical Reading and Writing

2. Do students with advanced, proficient, basic, and below-basic reading proficiency levels all benefit from the year-long instruction?
3. How does teachers’ fidelity of implementation with the core components of cognitive apprenticeship relate to student learning?

Setting and Participants

We recruited a total of 13 eighth-grade teachers in a large school district (in the top 25 in the nation in size) near a dense population center in the mid-Atlantic region of the United States, focusing primarily on schools within the district that traditionally served large numbers of struggling readers (about 33% of all students were significantly below grade level in reading) but also supporting the learning of more successful students (45% of students were proficient and just over 20% were advanced readers). The district serves a socially and ethnically diverse group of students, with 45% of the students receiving free or reduced-price meals, 8.5% receiving English services for speakers of other languages, and 10% receiving services for identified disabilities, according to federal criteria. In terms of demographics, during the year in which this study took place, district information indicated that African Americans made up the majority of the population at 74%, followed by students who were Hispanic at 17%, with relatively fewer Caucasian (5%), Asian (3%), and Native American (0.4%) students.

We enlisted teacher participants from different schools each year of the project. In this study, we recruited teachers from six different schools for the project, providing stipends and graduate credits in return for participating in PD in the intervention group. In sum, eight intervention and five control teachers agreed to allow their students to complete our pretests and posttests (two control teachers decided to use the curriculum intervention in subsequent years). Although intervention teachers were self-selected, control teachers had comparable education and experience in the profession. And while we did not collect data on the instruction that control teachers provided to their students, we can confirm that the participating school district expected all teachers to incorporate the teaching of reading and writing with primary sources. In addition, we observed intervention teachers each lesson (once per day), and we used fidelity of implementation data from the 144 resulting observations to inform our PD efforts and to learn more about underlying aspects of the curriculum intervention.

Approximately 1,330 students completed both pretests and posttests across the 13 teachers in the two conditions. After data collection ended, we selected a representative subset of students from teachers in each condition, who were stratified by reading proficiency level on the state-administered high stakes reading assessment, resulting in approximately 30% each of advanced, proficient, and basic readers, and 10% who we considered as below basic in terms
of reading proficiency because these students participated in an alternative assessment that was designed for students with significant cognitive disabilities who were unable to participate in the regular assessment. Although the state where this study occurred does not identify groups of below-basic readers, other states do, leading us to create a new category so we could evaluate our approach with students who struggled most with literacy as well as the most advanced learners. We decided to limit analyses of intervention data to students who completed at least five of six 3-day lessons. From this group, we randomly selected students for close analysis, being sure to select comparable numbers of students from different reading levels. The final pool included 153 intervention and 157 control students. Data from these 310 students were used for quantitative analysis.

Design

We used a mixed-methods design to address our research questions (Creswell, Clark, Gutmann, & Hanson, 2003), collecting quantitative and qualitative data concurrently in a nested design. Pretest–posttest data on student writing across intervention and control conditions allowed us to examine possible effects of the curriculum on students’ writing outcomes. Qualitative data from observer field notes, fidelity of implementation protocols, and teachers’ reflections about lessons provided insight into how teachers and students adhered to and adapted the curriculum. Some qualitative data, in the form of observer reports, were transformed into quantitative measures of intervention fidelity, allowing us to identify high- and low-fidelity teachers and explore the relationship between core elements to the intervention design and student learning outcomes (our third research question). Results from the combined analysis of qualitative and quantitative analyses were also subsequently used to refine the curriculum intervention and PD for the second year of our project.

Curriculum Intervention

The social studies curriculum in the cooperating school district focused on United States History in the eighth grade, with content standards that began with the Revolution and ended with Reconstruction. We chose six topics for our study in collaboration with district personnel. Initial negotiations with the district led to an agreement that teachers and students would be available for 18 days of instruction over the course of the school year. We worked with the district to identify each topic and then we planned a 3-day lesson sequence, or investigation, for each topic (see Table 1). The last investigation was different in that it combined two topics in three days. Each investigation included reading,
Table 1. Overview of Historical Investigations and Disciplinary Reading and Writing Strategies

| Historical investigation | Disciplinary strategies introduced
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Lexington Green</td>
<td><strong>Historical reading:</strong> Sourcing and contextualizing primary sources</td>
</tr>
<tr>
<td>Who fired the first shot at Lexington Green?</td>
<td></td>
</tr>
<tr>
<td>2 Shays’ Rebellion</td>
<td><strong>Historical writing:</strong> Identifying the components and structure of a historical argument</td>
</tr>
<tr>
<td>Were Daniel Shays and his followers rebels or freedom fighters?</td>
<td></td>
</tr>
<tr>
<td>3 Democratic-Republican Societies</td>
<td><strong>Historical writing:</strong> Planning an essay</td>
</tr>
<tr>
<td>Were Democratic–Republican Societies patriotic or subversive?</td>
<td></td>
</tr>
<tr>
<td>4 Indian Removal</td>
<td><strong>Historical reading:</strong> Discussing and evaluating evidence</td>
</tr>
<tr>
<td>What path offered the best chance of survival for the Cherokee in the early 1800s: staying in their original territory or removal to the west?</td>
<td><strong>Historical writing:</strong> Composing a full essay</td>
</tr>
<tr>
<td>5 Abolitionism</td>
<td>Students set goals to read, analyze, plan, and compose with greater independence</td>
</tr>
<tr>
<td>What was the most promising path toward freeing slaves in the U.S. before the Civil War: nonviolence (“moral persuasion”) or more aggressive action?</td>
<td></td>
</tr>
<tr>
<td>6A Texas Annexation</td>
<td>Students integrate reading analysis, planning, and composing independently</td>
</tr>
<tr>
<td>Was the annexation of Texas good for the United States in 1845?</td>
<td></td>
</tr>
<tr>
<td>6B Mexican American War</td>
<td></td>
</tr>
</tbody>
</table>
We listed a short question series on a handout to help students both understand and internalize discrete concepts of perspective, contextualization, and substantiation as part of a disciplinary approach to reading. For example, we asked students to consider “What is the author’s point of view or position?” and “Why did the author write this document?” to facilitate thinking about perspective. To promote contextualization, we asked students to consider the type of document, e.g., “What type of document is this and where did it appear?” and “What was the occasion for which this was written?” as well as prompting students to consider the time period and setting in which it was written or spoken, e.g., “What else was going on at the time?” and “How long after the event was this written?” and so on. Further, we attempted to facilitate corroboration by asking students to determine whether documents agreed or disagreed with each other and, finally, through these acts, to weigh the evidence and construct an interpretation to promote thinking about evidence and substantiation.

We helped students to simultaneously focus on conveying their interpretations in written arguments through a different set of scaffolds that were intended to build students’ background knowledge about the writing process and about historical argumentation. First, we provided and helped students to examine sets of sample essays that were written with a predictable text structure (e.g., providing an introduction that explained the historical controversy and gave a response to the historical question and providing a conclusion that compared both documents and explained why evidence that was selected was more convincing than other evidence in the documents), taking care to provide different but equally compelling evidence in response to the controversy. Students compared and contrasted sample essays, discussing the historical ideas and the text structure, then brainstormed how to improve the sample essays by engaging in the above historical reading process with the primary source documents.

Our final goal was to help students to learn that planning required a distinct set of actions from composing. To do this, we asked students to reconcile opposing responses to the historical questions through a new set of prompts that focused on substantiation (e.g., identify a reason for one side, select evidence that supports this reason, explain how the evidence supports the reason, and evaluate the reason; and then to engage in the same process for the opposing view). Students made notes about their response to these prompts as part of written plans. We hoped this process enabled students to see the importance of planning, as well as structure, their attempts in selecting, explaining, and evaluating evidence that they thought was most convincing in their arguments.

During the first half of the year (Investigations 1 through 3), teachers modeled application of these heuristics. The rest of the year (Investigations 4 through 6) was primarily focused on collaborative and independent practice in applying them, with an increasing focus on how students were to manage the reading and writing processes on their own. Teachers frequently asked students
to reflect on what they were learning and how underlying components of the intervention related to and supported the overall goal of writing argumentative essays.

**PD**

In addition to having constraints in the number of days in which to unfold our curriculum, we also worked within the district’s framework for conducting our PD, which meant holding full-day sessions on a limited number of Saturdays. As a result, we conducted four workshops to (a) establish a shared view on disciplinary reading and writing in history and with adolescent learners; (b) demonstrate how to use the specific supports and model of instruction in our intervention; and (c) encourage reflection on what students were learning, how to respond to challenges they were seeing with the curriculum, and their role in teaching the lessons. The sessions began with activities (e.g., comparing textbook with primary accounts) that gave teachers an understanding of historical concepts, such as evidence, accounts, perspective, and context, as well as knowledge of historical reading, deliberation, and writing. Then, we introduced teachers to the curriculum, with one or more investigations per meeting, and used these investigations as opportunities to delve more deeply into the concepts introduced during the initial sessions.

In each PD session, we modeled the use of the curriculum materials, debriefed the key elements of each investigation, talked through how teachers might enact these elements, and gave teachers opportunities to practice teaching key aspects of the investigation to their peers. Thus, we asked teachers to apply the mnemonics and scaffolds first as learners, and then gave them opportunities to reflect on the use of scaffolds with students. Teachers were also given time to debrief about lessons and share what was working with each other and to draw on successes from the collaborative workshop experience. Finally, because we reviewed lesson materials from several investigations during each meeting, we also met with teachers during their planning time during or after school, to deliver “just in time” reminders of the goals for each lesson and essential components.

**Writing Task**

We asked students to compose historical arguments using two primary sources in response to one central historical question, “Were African Americans free after the Civil War?” at both pretest and posttest. Students had not learned about the post-Civil War era before either test. We created two forms of this test (see Appendix A) to allow us to counterbalance the measures. Both forms asked the same question, but each used a different document set.
Form A included an excerpt from an article in *The Atlantic Monthly*, written in May and June 1864, entitled “Life on the Sea Islands.” In this letter, Charlotte Forten, an African American from the north, described her experience at the end of the Civil War in teaching newly freed slaves. The second document was an excerpt that had been adapted from a letter written by Captain C. M. Hamilton in 1866 to the Office of the Adjutant General in Washington, DC. The officer was a northerner in the U.S. Army who was posted in Florida after the Civil War, and his letter describes events at a school in the region where he was posted.

Form B included two documents that provided students with information about African Americans’ lives and opportunities to pursue individual freedom. Documents in each set were paired to contrast the positive and negative experiences of African Americans during Reconstruction. We counterbalanced the presentation of these tests so that some students were randomly assigned to respond to Form A at pretest and others were randomly assigned to respond to Form B at pretest. We then switched which form students responded to at posttest. In this way, we minimized the impact of the tests on the results we found.

Researchers in history education have used similar tasks to assess students’ historical thinking and writing (Rouet et al., 1996; Seixas, 2006; Young & Leinhardt, 1998). Our instrument is consistent with notions regarding analysis of evidence, use of evidence to construct interpretations of the past, and communication of arguments in writing. While such practices echo the work of historians, they differ in that historians typically come up with their own questions and discover evidence through archival research. Obviously the nature of an in-class test does not allow for such practices.

To ensure that the tests were appropriate for students’ age and literacy levels, we made several changes to the primary sources (Wineburg & Martin, 2009). We excerpted them, focusing on segments that were most relevant to the question, so that the sources were no more than one page. We created a head note at the beginning of the source to orient readers to the texts and offer background knowledge that might help them make sense of the texts. We inserted an attribution at the bottom of the source to give students information, such as the date, place, genre, and author of the text, to allow for a historical reading of the text. Finally, we substituted simpler vocabulary or phrasing where necessary to attain Lexile scores appropriate for sixth-graders on both tests since several participating students were two or more years below grade level in reading skills.

**Data Sources**

We analyzed students’ historical essays using three writing measures, focusing on their ability to write historical arguments, the overall quality of
their writing, and the length of their essays both before and after the year-long historical thinking curriculum intervention. We analyzed teachers’ scores from fidelity of implementation protocols to examine constructs that we believed were underlying effective implementation of the curriculum intervention, then examined observer field notes, lesson summaries from fidelity protocols, and teachers’ reflections about lessons to explore strengths and weaknesses in our approach to PD.

**Historical arguments.** This dependent variable served as a measure of specific aspects of historical thinking and was based on an analytic trait rubric that focused on three specific aspects of historical reasoning—substantiation, perspective recognition, and contextualization—resulting in a separate score for each. Substantiation emphasized the extent to which students provided evidence and explanation in support of a claim. Perspective recognition focused on students’ skills in presenting the texts as authors’ viewpoints rather than as authoritative words to be accepted literally. Contextualization addressed the extent to which students identified and situated their argument and primary sources in the appropriate time, place, and setting, thus linking related events.

We taught two pairs of raters to use the analytic trait rubric, asking them to consider one trait at a time and to talk through distinctions in scores. The raters scored the entire set of 620 pretest and posttest essays, working in pairs, and achieved satisfactory reliability for each analytic trait (for substantiation, Pearson $r = .87$; for perspective, $r = .93$; for contextualization, $r = .88$). The separate scores were combined, and the summary score was standardized to have a mean ($M$) of 0 and a standard deviation ($SD$) of 1. Pretest historical thinking scores ranged from –1.80 to 2.35 in the full sample and from –1.74 to 2.30 in the treatment group sample. Posttest scores ranged from –1.66 to 2.26 in the full sample and from –1.78 to 1.82 in the treatment sample.

**Holistic quality.** This measure assessed the clarity and persuasiveness of students’ responses to the historical question, basing scores on a holistic rubric (with ratings from 0 to 6). The highest score was awarded to papers with a clear, purposeful essay that was both persuasive and well structured, and the lowest score was assigned to papers that ignored or misunderstood the prompt. We included this measure as a global indicator of writing and intended it as a check as to whether our intervention, which focused on improving students’ use of claims and evidence, would be sufficient to improve their generic ability to write clear and persuasive essays or if we would need to explore whether students needed more direct intervention in writing or more time devoted to writing for such improvements in future years of the project.

As an example of a paper between these ratings, a paper awarded a 4 was judged to be clear but with little development in persuasiveness or structure. We taught three different pairs of raters to use the holistic rubric to avoid potential crossover effects associated with asking the same readers to score essays
for more than one dependent measure. The raters scored the complete set of essays with 93% exact agreement. The measure was standardized to have a mean (M) of 0 and a standard deviation (SD) of 1. Pretest clarity and coherence scores ranged from –2.59 to 2.78 in the full sample and from –1.61 to 2.74 in the treatment sample. Posttest scores ranged from –2.26 to 2.17 in the full sample and from –1.89 to 1.90 in the treatment sample.

**Essay length.** This dependent variable consisted of the number of words written, regardless of spelling, as a way to capture students’ degree of fluency in generating written text. We expected that length of composition would be a reasonable gauge of overall writing facility, because prior studies on the relationship between word count and essay quality have found there to be a significant positive relationship between the two, with average correlations ranging from the .60s (Kobrin, Deng, & Shaw, 2007) to a near-perfect correlation when students use word processors (r = .99; Bangert-Drowns, 1993). Although students may compose long papers that are of poor quality, length can serve as a useful proxy for more global skill, since a reasonably well-written longer composition requires the use and coordination of a variety of underlying processes (handwriting, text generation, motivation, and so on; Morphy & Graham, 2011). Agreed upon scoring conventions included counting “nine o’clock p.m.” as three words, “Military” as one word and “United States” as two words, and “1863–1865” as three words. Independent raters scored all essays. A random sample of 50 papers was counted twice with adequate reliability (Pearson r = .99). This measure was standardized to have a mean (M) of 0 and a standard deviation (SD) of 1. Pretest essay length ranged from –2.02 to 3.89 in the full sample and from –1.98 to 3.83 in the treatment sample. Posttest scores ranged from –1.74 to 4.01 in the full sample and from –1.70 to 3.08 in the treatment sample.

**How teachers used the curriculum.** To determine whether and how teachers implemented our historical thinking curriculum, we developed observation protocols for coding constructs of the intervention that we asked teachers to implement (see Appendix B for a sample protocol for the first day of Investigation 1). Results from this tool then helped us evaluate the effects of differing levels of fidelity. Scores represented the percentage that the observer could identify important constructs that were part of each lesson.

When presenting the intervention to teachers, we highlighted these constructs, informing teachers that these elements were critical, while also giving teachers freedom, when necessary, to implement them in ways that they thought made the most sense for students. We mapped each critical element to one of four constructs of the intervention, based on principles of cognitive apprenticeship, which we believed to be instrumental in helping students gain independence in their learning. The four constructs follow.
Developing Historical Reading and Writing

1. **Building understanding of strategies.** This construct includes whether the teacher modeled the historical reading and writing strategies, described critical features of what they have modeled, reviewed the “how to” of the strategies, and established the purpose of each strategy.

2. **Promoting independence in using strategies.** This construct was a measure of the extent to which the teacher established the purpose for each lesson, checked for understanding of the strategies, supported students by providing guidance and feedback, and promoted autonomy and students’ understanding of strategies.

3. **Building historical and disciplinary knowledge.** This construct was a measure of the extent to which the teacher provided background knowledge, content, and vocabulary; explained the historical controversy; checked for historical understanding of primary source documents; and understood the historical content.

4. **Classroom management.** While this construct is not particular to cognitive apprenticeships, we included it to measure the impact of learning environment on teachers’ and students’ ability to engage in the key features of cognitive apprenticeships. This variable indicates whether the classroom was organized in a way that allowed students to focus and follow the lesson and the degree to which students and teachers showed respect for each other.

We averaged scores across lessons to estimate a fidelity score for teachers who participated in the intervention, then conducted a principal components factor analysis using all four fidelity constructs to create a single fidelity factor score. The resulting factor scale had an eigenvalue of 3.71 and explained 92.7% of the combined variance. Cronbach’s alpha for the scale is .97. By looking at the degree to which teachers accurately implement the core components of an intervention, we get more data on whether these components are associated with learning outcomes (O’Donnell, 2008). Second, fidelity data provide useful insights into the challenges that present themselves as teachers put interventions into practice (Century, Rudnick, & Freeman, 2010). Understanding these challenges is critical to refining intervention design for future implementation and designing PD.

Our goals were to assess whether teachers’ level of fidelity influenced students’ learning outcomes and to determine the relative importance of each of these constructs in the success of the intervention and to consider the rationale teachers gave for any departures. The first three of these constructs were developed to measure the degree to which the core components of the intervention were implemented as intended, or treatment delivery (Schulte, Easton, & Parker, 2009), while the fourth measured the degree to which the teacher established an environment in which students could engage in the intervention activities, or treatment receipt (Schulte et al., 2009).
**Student work completion.** We also explored the extent to which students completed lesson components that corresponded with disciplinary activities (reading, evaluating, planning, and writing), following O’Donnell’s (2008) suggestion that researchers examine students’ adherence to the structural components of an intervention. That is, even if teachers were faithful in their lessons, students might be absent or fail to finish critical components of the intervention.

**Statistical Methods and Analytic Measures**

We analyzed our data using hierarchical linear models (HLMs) for all analyses because students are nested within classes. HLM is the preferred approach because the data are clustered—students within classrooms—and statistical approaches that ignore this clustering result in standard errors that are too small, which contributes to errors of inference and compromises the validity of results. Although students in our study were grouped within classes, and classes were grouped within teachers, the small number of teachers limited the statistical power of a three-level model (with students within classes, within teachers), so we modeled the nesting of students within classes. Analyses with students nested within teachers yielded similar coefficients that were not statistically significant, suggesting a power issue rather than substantive differences between modeling within classes and within teachers.

Modeling students within classrooms allows us to focus on student learning outcomes while taking into account incoming characteristics, such as students’ race and academic background, and between-class variability in the outcomes using variables associated with the class, such as exposure to the intervention (Arnold, 1992). We fit three separate models, one for each outcome: elements of reasoning in students’ historical arguments, the holistic quality of their writing, and the total length of each essay. HLM allows researchers to “model the relationships between several predictors and the dependent variables simultaneously so that the results indicate the unique effect of each predictor when accounting for other predictors” (Bowman, 2012, p. 379), giving a more complete picture of the results across variables.

Before testing hypotheses about differences between intervention and control conditions, we examined differences between the intervention and control schools at pretest using analysis of variance. This analysis was undertaken to ensure comparability of the groups before instruction began because students were not randomly assigned to conditions and because teachers volunteered to be in each group (intervention and control). We then estimated unconditional models (results reported below) to assess the magnitude of variance attributed to classes and students within classes as a check on the reasonableness of our decision to use HLM.
Effects of the curriculum intervention on students’ posttest scores were estimated, controlling for students’ pretest scores on each measure. Outcomes were also estimated, controlling for students’ gender (the referent group for the analysis is female students), ethnicity (the referent group is African American students), and reading proficiency as measured by the state-mandated assessment from the prior year (the referent group consists of students who scored at the basic level of proficiency in reading or who took an alternate assessment designed for students with significant cognitive disabilities), all of which were entered into the multilevel models for posttests at the student level. Each student-level covariate was centered at the grand mean.

**Fidelity Outcomes**

Our intervention variables examine the extent to which students received the intervention. In preliminary analyses, we noticed that teachers who were delivering curriculum fell into two groups: those who implemented the curriculum with relatively high fidelity and those who implemented the curriculum with less fidelity. To explore the effects of this difference, we created two dichotomous variables that indicated whether teachers’ level of participation in the intervention demonstrated high or low fidelity to the curriculum intervention. When the teacher’s fidelity score was greater than the mean, the class was considered to be in the high treatment condition. Teachers who had a score on teacher fidelity equal to or less than the mean were considered to be in the low treatment condition. There were roughly equal proportions of students taught by high- and low-fidelity teachers within the treatment condition. When high treatment and low treatment are both in the model, the referent group is the control group.

**Student Work Completion**

We developed coding procedures to tabulate each student’s success in completing critical lesson components. Graders were given an extensive protocol for all lessons and recorded students’ work completion, which was then used to create an overall student work completion fidelity score. We calculated fidelity scores for each student and then averaged these scores across all students in each class. These averages provide information regarding the overall level of student fidelity to the curriculum intervention within a given class. The aggregate student work measure might be viewed as an indicator of the teacher’s success at classroom management (i.e., the extent to which the teacher established a learning environment that allowed students to focus and follow the lesson). Aggregate student work and classroom management are quite highly correlated ($r = .89, p < .001$). This high correlation
between student fidelity and classroom management provides some additional validation of our observation protocol, at least in terms of classroom management.

**Statistical Models**

We used a series of two-level random intercept models, with students at Level 1 and classes at Level 2, to examine the effects of participating in the curriculum intervention on three aspects of students’ writing skills: historical arguments, holistic quality, and essay length. We estimated these models using restricted maximum-likelihood estimation, the preferred estimation strategy for models with relatively few level-2 units (McCoach, 2010). Student-level variables included gender, a series of dummy variables indicating racial status, pre-intervention proficiency on the state reading assessment, and pre-intervention scores for the respective outcomes (historical arguments, holistic quality, and essay length).

At Level 2, we model the average writing skills of students as a function of participation in the curriculum intervention and random classroom error. In analyses that focused on the importance of different dimensions of fidelity, we restricted our analysis to the classes and students participating in the curriculum intervention. The Level-1 model is the same as above, but at Level 2, we substitute a measure of fidelity for the dummy-coded treatment variable. We model the average writing skills of students as a function of fidelity to the curriculum intervention and random classroom error. Our models allow the intercept for post-intervention writing skills to vary randomly between classrooms. Because all student-level variables were grand-mean centered, the Level-1 intercept is the average writing skills of students in classes with average student characteristics, including average percentages of students who were viewed as proficient or advanced on the state reading assessment and average pretest scores on the dependent measures.

**RESULTS**

**Descriptive Results**

Table 2 contains summaries of the student variables, and Appendixes C and D provide written responses to pretests and posttests from a reader with below-basic proficiency and advanced proficiency. We computed t-tests and chi-squared analyses to examine the extent to which the treatment and control groups had comparable student characteristics and comparable pretest scores for the disciplinary writing assessments. The only statistically significant demographic differences \(p < 0.05\) were that the intervention group had...
Table 2. Descriptive Statistics for Full Sample, Control Group, and Treatment Group

<table>
<thead>
<tr>
<th>Variables</th>
<th>Full sample, N = 310</th>
<th>Control group, N = 153</th>
<th>Intervention group, N = 157</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male</td>
<td>51.3%</td>
<td>52.9%</td>
<td>49.7%</td>
</tr>
<tr>
<td>Black</td>
<td>56.5%</td>
<td>67.3%</td>
<td>45.9%</td>
</tr>
<tr>
<td>White</td>
<td>2.9%</td>
<td>4.6%</td>
<td>1.3%</td>
</tr>
<tr>
<td>Hispanic</td>
<td>37.4%</td>
<td>24.8%</td>
<td>49.7%</td>
</tr>
<tr>
<td>Asian/Pacific Islander</td>
<td>3.2%</td>
<td>3.3%</td>
<td>3.2%</td>
</tr>
<tr>
<td>Proficient reader</td>
<td>30.3%</td>
<td>29.4%</td>
<td>31.2%</td>
</tr>
<tr>
<td>Advanced reader</td>
<td>30.0%</td>
<td>30.7%</td>
<td>29.3%</td>
</tr>
<tr>
<td>Pre-historical argument</td>
<td>0.00 (1.00)</td>
<td>-0.00 (0.97)</td>
<td>0.01 (1.03)</td>
</tr>
<tr>
<td>Pre-holistic quality</td>
<td>0.00 (1.00)</td>
<td>0.04 (0.97)</td>
<td>-0.04 (1.03)</td>
</tr>
<tr>
<td>Pre-essay length</td>
<td>0.00 (1.00)</td>
<td>0.00 (0.99)</td>
<td>-0.00 (1.02)</td>
</tr>
<tr>
<td>Post-historical argument</td>
<td>0.00 (1.00)</td>
<td>-0.28 (0.81)</td>
<td>0.27 (1.09)</td>
</tr>
<tr>
<td>Post-holistic quality</td>
<td>0.00 (1.00)</td>
<td>0.05 (0.79)</td>
<td>-0.05 (1.17)</td>
</tr>
<tr>
<td>Post-essay length</td>
<td>0.00 (1.00)</td>
<td>-0.30 (0.60)</td>
<td>0.30 (1.20)</td>
</tr>
</tbody>
</table>

*p < .05, ***p < .001.

A greater proportion of Hispanic students (49.7% versus 24.8% in the control group) and fewer African American students (45.9% versus 67.3% in the control group). There were no significant differences between groups on beginning of the year proficiency levels or pretest scores on the writing outcomes of interest (*p < 0.05), meaning that groups were comparable before instruction began with respect to students’ historical arguments, holistic quality, and essay length. The posttest scores, however, were significantly different. Differences favored students in the curriculum intervention group for historical arguments (*ES = .56) and essay length (*ES = .60), while the control group slightly outperformed the intervention group on our measure of holistic quality (*ES = .10).

Multilevel Models

We first fit fully unconditional models to examine the variability between the classes in the three end-of-year writing scores. The intraclass correlation, which is the proportion of the total variation in the dependent variable between classes, was .42 for the posttest historical argument score, .38 for the posttest holistic quality, and .44 for the posttest essay length. These results indicate that a multilevel approach is needed to produce accurate estimates of student learning outcomes and to avoid overestimating the significance of the intervention. We then fit a Level-1 model for predicting students’ writing outcomes as a
function of their demographic characteristics and a Level-2 model to determine the effects of the curriculum while controlling for other variables. For the sake of simplicity, we present only the results of the fully conditional model in Table 3.

**Curriculum Effects**

Table 3 presents the two-level fully conditional models for scores on historical arguments, holistic quality, and essay length. The second column presents the results for historical arguments, the third column presents results for holistic quality, and the fourth column presents results for essay length. Because the dependent variables are standardized, coefficients can be interpreted as effect sizes, which give a sense of the magnitude of the findings. Cohen (1988) hesitantly defined effect sizes as “small, $d = .2$,” “medium, $d = .5$,” and “large, $d = .8$” (p. 25).

The results in Table 3 indicate that the intervention had positive effects on both post-historical arguments and essay length but not holistic quality, and that the size of the effect is related to the extent to which teachers implemented the program with fidelity. These results represent curriculum effects after controlling for gender composition, racial and ethnic background, and multiple measures of prior achievement. Students in classes led by teachers

### Table 3. Estimated Effects of Disciplinary Learning in Writing Intervention on Disciplinary Writing Skills

<table>
<thead>
<tr>
<th>Variables</th>
<th>Post-historical argument</th>
<th>Post-holistic quality</th>
<th>Post-essay length</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>-.274</td>
<td>0.032</td>
<td>-.287</td>
</tr>
<tr>
<td>High treatment</td>
<td>.648***</td>
<td>-0.076</td>
<td>.602**</td>
</tr>
<tr>
<td>Low treatment</td>
<td>.322*</td>
<td>-0.249</td>
<td>.441*</td>
</tr>
<tr>
<td>Proficient reader</td>
<td>.357**</td>
<td>0.501***</td>
<td>.278*</td>
</tr>
<tr>
<td>Advanced reader</td>
<td>.815***</td>
<td>1.052***</td>
<td>.623***</td>
</tr>
<tr>
<td>Pre-test</td>
<td>.031</td>
<td>0.097*</td>
<td>.225***</td>
</tr>
<tr>
<td>Variance component</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept</td>
<td>.176***</td>
<td>0.125***</td>
<td>.189***</td>
</tr>
<tr>
<td>Reliability</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept, B0</td>
<td>.623</td>
<td>0.541</td>
<td>.653</td>
</tr>
</tbody>
</table>

*p < .10, *p < .05, **p < .01, ***p < .001.
Developing Historical Reading and Writing

with high fidelity to the curriculum intervention outscored students in control classes on the post-historical arguments ($ES = .65$) and on essay length ($ES = .60$). Thus, for these two outcomes, the intervention had a moderate to strong effect on students whose teachers had high fidelity to the intervention. Students whose teachers implemented the curriculum intervention with less fidelity still outscored students in the control group, though less dramatically: $ES = .32$ on the measure of post-historical arguments and $ES = .44$ on the essay length measure. There were no differences between students in holistic quality.

The analyses also indicate moderate to large effects associated with differences in students’ individual learning characteristics with the state-administered test of reading proficiency. Students with advanced reading proficiency substantially outscored students with basic and below-basic proficiency, with differences ranging from $ES = .62$ for essay length and $ES = 1.05$ on the measure of holistic quality. Students with proficient levels of reading also outscored students who earned basic levels of proficiency in reading. Although the magnitudes of the differences are smaller, they still range from one-fourth ($ES = .28$ on essay length) to one-half of a standard deviation ($ES = .50$ on holistic quality).

Although students with higher levels of proficiency outscored students with lower levels of proficiency, this does not mean that they benefited more than other students from being exposed to the historical thinking curriculum intervention. All students benefitted in that their posttest scores show statistically significant improvement from their pretest scores. However, students who began with the highest levels of literacy showed the most advanced historical thinking at the end of the year on average. To examine whether the curriculum had differential effects for students with different proficiency levels, we modeled the slope for the two dummy-coded variables for proficiency by whether students received the curriculum or not. This analysis provided no evidence that the curriculum effect was greater for students with higher levels of proficiency compared to students with lower levels of proficiency. To further test the possibility of differential effects, we calculated the proportion of students with basic, proficient, or advanced reading scores for each class and then examined a set of interaction terms at the class level. Again we found no evidence that the curriculum effect for classes with greater proportions of lower-proficiency students were any different from the curriculum effects for classes with greater proportions of higher-proficiency students—students, regardless of proficiency level, benefited equally from exposure to the historical thinking curriculum intervention. Sample pretest and posttest essays, from one student who was below basic reading proficiency and one advanced reader, are presented in Appendixes C and D.

To get a better sense of the ability of these models to predict the outcomes, we calculated the proportion of variance explained. The fully conditional models in Table 3 explain .58, .68, and .56 of the total variance for the writing measures (historical arguments, holistic quality, and essay length,
respectively). Each of these numbers can be interpreted as an estimate for \( r^2 \) associated with the full models (which include student demographics and academic background as well as participation in the intervention) relative to the unconditional models. When we explore the proportion of variance explained by the intervention alone, we find that the curriculum effect = .27 of the between-class variance for the measure of historical thinking and = .28 for the length of students’ essays.

Examples of student writing are shown in Table 4, which also provides an overview of our rubric for historical argumentation. This table provides criteria for substantiation, perspective recognition, and contextualization ratings from 0 to 3. Because 310 students’ papers were analyzed before and after the curriculum intervention, the excerpts shown here should be viewed only as illustrations, with many more variations in how students demonstrated their understandings. Moreover, these examples come from both struggling and advanced readers—some of the most advanced readers received poor ratings at pretest, and some struggling readers received the highest ratings at posttest.

Fidelity Effects

In addition to examining the overall impact of the curriculum intervention across groups, we also wished to learn which parts of our lessons were important to the intended writing outcomes for students who participated in the historical thinking curriculum. As described earlier, we identified four constructs critical to helping students learn to write historical essays, and for the statistical analyses, we standardized each measure across teachers in the curriculum intervention condition (thus each mean is 0 and has a standard deviation of 1).

Therefore, we ran a series of models with different Level-2 predictors within the curriculum intervention group. The separate models examine the potential impact of five different measures of fidelity on students’ end-of-year assessments of writing skills. The first model examines only student fidelity (average work completed per class), and the second through fifth models examine the individual dimensions of teacher fidelity (content, independence, modeling, and management). All of the student-level variables were grand-mean centered, and all continuous measures are standardized, including the dependent variables (\( M = 0, SD =1 \)). Coefficients can be interpreted as effect sizes.

Figure 1 summarizes the results for all three writing outcomes for the five different models of fidelity. Looking across the bar graph, fidelity has the strongest effects on historical arguments (darkest-shaded bars). Each of the measures of fidelity constructs had statistically significant and positive relationships with the outcomes. The largest effects are associated
<table>
<thead>
<tr>
<th>Substantiation</th>
<th>Perspective recognition</th>
<th>Contextualization</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 (a) No evidence. OR (b) Irrelevant evidence. OR (c) Transcription of document (and nothing else).</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(a) Presents evidence from documents as student’s own perspective. OR (b) Treats documents as authoritative (e.g., “Document 1 says . . .” “It says in the document . . .”).</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(a) No context mentioned. OR (b) Student uses anachronisms (e.g., makes a chronological mistake or uses information from another time period w/o noting the different era).</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note: There may be a claim

“Yes, African Americans were free after the civil war because in the civil war the north fought for them to be free. If they were not free they would not be going to school. They were not entirely free because people still treated them wrong.”

“because in Document two”

“The North did not want slavery but the South did so that is why the south created laws that stopped them from voting, eating with other people, drinking from the same water fountain, and riding on the bus where they wanted to sit.”

1 (a) Facts/quotes presented without explanation (may implicitly support a conclusion that is stated). OR (b) Facts/quotes are presented with inaccurate explanations throughout.

“Captain Hamilton on the other hand thought differently. He thought it was terrible. He states in his letter ‘one evening a mob called the teacher out of the school house and the teacher came face to face with four revolvers. From my point of view that school was just unorganized and don’t know how to control their students.’

(a) Mentions the author(s) (e.g., “According to Lynch . . .” “The author says . . .”).

(Continued)
<table>
<thead>
<tr>
<th>Substantiation</th>
<th>Perspective recognition</th>
<th>Contextualization</th>
</tr>
</thead>
</table>
| 2  
(a) Facts/quotes are presented without explanation, but are explicitly linked to a conclusion. 
OR  
(b) Facts/quotes are explained but not explicitly linked to a conclusion. AND At least one explanation is accurate. | (a) Describes author’s perspective in a way that recognizes text as the author’s point of view.  
*Note: The author’s name does not have to be mentioned* | (a) Describes contextual details alongside the evidence without explicitly connecting them. |

"I think that African Americans were not freed after the Civil War. Because in ‘Document 2’ it says that they had ‘houses broken open, windows smashed and doors broken down in the dead of the night.’ Also it says..."  

"After I read both letters I think that African Americans were never free after the Civil War. In the letter it said, ‘But when we are attacked at the midnight hour, our lives threatened and the laws fail to protect or help us, the only thing we can do is defend ourselves.’ They are scared and threatened by these people who only wish to harm them. When you’re free you aren’t scared of anything or anyone like these people are.”  

"This is happening in Baltimore & Georgia—when 1865 to 1867..."
(a) Facts/quotes are explained AND linked explicitly to a conclusion.  
AND  
(b) Explanations are consistently accurate.  

(a) Evaluates the authors’ perspectives (e.g., discusses reliability/trustworthiness).  
OR  
(b) Reconciles multiple authors’ perspectives (e.g., compares the ideas in the documents).  

(a) Integrates context and evidence in an explanation or conclusion.  
OR  
(b) Uses context and evidence together to draw a conclusion or make an inference.  

... In document 2 African Americans considered themselves not free. They say they are threatened. Some of the people who agree say that “we wish to live in peace and quiet. But when we are attacked at midnight hour, our lives threatened and the laws fail to protect or help us, the only thing we can do is defend ourselves.” They think that even though they’re supposed to be free, they are still threatened for little things. I think they mean that they aren’t really free if they are attacked for nothing.  

... In my opinion I don’t think they are free. I support this opinion because of the quote stated in paragraph 3 of this essay taken from document 2. This convinced me because even though some African Americans can do more than they used to, some of them can’t do the same. Although they have the same rights as others. So I agree that document 2 is more convincing.  

... Twenty four African Americans from Calhoun, Georgia wrote a letter to the commander of the third military District on August 25, 1867 and describe their complaints for protection. This shows a different point of view which was that the slaves were not free after the Civil war...  

... The 24 African Americans say that they “wish to do right, obey the laws and live in peace and quiet” but they ask for protection. This shows that even though they receive buildings being treated right means more to them...  

... James Lynch an African American minister from Baltimore who worked as a missionary and teacher in Savannah, Georgia during Reconstruction, wrote a to the New York’s Freedman’s Relief association on January 4, 1865. His point of view was that the slaves were freed after the civil war. In his letter he states that the colored people did not seem to realize that they were free. It was not announced to them. He says “I am happy to say that their confidence has been growing every day and they seem to lose that fear which slavery had made a second nature.” This shows that the colored people feel like they are part of the nation now, after the civil war. The government gave them three large buildings they will use for school purposes and one as a hospital...
with classroom management ($ES = .48$) and the student work completion variable ($ES = .45$). These constructs are highly correlated, and our findings suggest that strong classroom management may support greater student work completion. Smaller, but positive, effects are associated with content ($ES = .40$), independence ($ES = .39$), and modeling ($ES = .29$).

Compared to historical arguments, the level of fidelity has noticeably smaller effects on the students’ holistic quality (lightest-shaded bars), which might be expected given that there was no curriculum effect for this measure. The largest effects are again associated with management (.25) and student fidelity (.29). Slightly smaller effects are associated with content (.21) and independence (.23). There is no effect associated with teachers’ modeling of strategies, possibly due to the overlap between classroom management and other aspects of fidelity. The measures of fidelity also have smaller effects on students’ ability to write longer essays (medium-shaded bars) than on historical arguments. The largest effects remain associated with management ($ES = .28$) and student fidelity ($ES = .29$), with slightly smaller effects associated with independence ($ES = .21$). There are no effects associated with the content or modeling dimensions of fidelity.

In sum, the fidelity measures have stronger associations with the quality of students’ historical arguments than with holistic quality and essay length, and their strongest associations were student work completion and teachers’ classroom management.

**DISCUSSION**

The results from this study add to growing literature on the positive impact of scaffolded instruction on middle and high school students’
Developing Historical Reading and Writing (e.g., De La Paz, 2005; De La Paz & Felton, 2010; Nokes et al., 2007; Reisman, 2012a). Although tentative by virtue of design limitations (i.e., we were not able to randomly assign teachers to conditions, did not have enough teachers to use them as the unit of analysis, and were unable to observe control classrooms), the writing data presented here come from and are representative of a large number of students (N = 620 essays) and provides support for the contention that when delivered with fidelity, our curriculum intervention and PD program resulted in improved historical writing for academically and culturally diverse learners, especially in comparison to writing from students who were in eighth-grade classes where the program was not provided or not delivered in ways that were faithful to the design of our lessons. Our results also indicate the success of our cognitive apprenticeship approach for readers at the highest proficiency levels, not only those who struggled academically. Finally, we learned how teachers’ actions influenced the impact of the curriculum intervention, leading us to make revisions both in the form of our curriculum intervention and in the way we provided PD in the second and third years of our project.

Writing Outcomes

We found that whereas at the beginning of the year students in both groups demonstrated similar abilities in writing, at the end of the year, there were clear and meaningful differences favoring the students who received the curriculum intervention in their ability to write historical arguments (equivalent effect sizes of .65 for students whose teachers implemented the curriculum intervention with high fidelity and .32 for students whose teachers implemented the curriculum with low fidelity) and in their ease in writing more content (.61 for students whose teachers implemented the curriculum intervention with high fidelity and .44 for students whose teachers implemented the curriculum intervention with low fidelity), after accounting for variation in students’ prior achievement, gender, ethnicity, and differences in these characteristics across classrooms. To be fair, however, we acknowledge that this early version of our curriculum intervention did not result in equal gains on a measure of students’ overall writing quality (i.e., the persuasiveness of their argument and its overall organization). These results, while somewhat disappointing, suggested that improvements in students’ use of claims and evidence were not sufficient to impact a more generic writing indicator. In reviewing these results, we thought it was possible that students who engaged in our curriculum intervention improved in their ability to write disciplinary arguments and in their general writing facility but needed more time in writing to show greater gains, prompting us to make changes to the curriculum intervention in later years of the project. These changes proved effective in enhancing students’ overall writing quality while maintaining improvements with their historical arguments and essay length (De La Paz et al., 2013).
Benefits for Learners With Varying Reading Profiles

Our results also indicate that within this group, students with higher levels of proficiency (and, at least in one instance, higher pre-intervention writing skills) outscored students with lower levels of proficiency and lower pretest assessment scores (see Appendixes C and D for samples from two such students—in fact, close examination of these papers reveals that the struggling reader ran out of time in composing her posttest essay after spending a significant portion of time in planning, actually resulting in decreased posttest scores for some variables). However, this does not mean that more advanced readers benefited more than less-proficient readers from being exposed to the disciplinary writing curriculum. Instead, our data showed that students benefitted equally from exposure to the disciplinary curriculum. In other words, although proficient and advanced readers began (and ended) with higher levels of writing proficiency, the students who were at the basic reading proficiency level made comparable learning gains (this result was also reported in De La Paz’s [2005] study with academically diverse students). While it would have been even more encouraging to “narrow the gap” in the struggling readers’ writing abilities, our data show that weaker readers did benefit from the curriculum intervention by growing at a similar rate as the more advanced readers.

Fidelity of Implementation

The degree to which teachers implemented the intervention mattered, and we found evidence that several underlying dimensions of the curriculum were central for student learning. However, the relationships among the underlying dimensions appeared more complex than we had hypothesized. The fidelity measures were most strongly associated with writing historical arguments, and we found weaker associations with the overall quality and essay length measures. Moreover, while overall fidelity and the degree that students completed critical lesson components were related to all writing outcomes, the effects of teachers’ content knowledge, modeling, and their ability to promote independence had less impact in comparison to the influence of classroom management. Classroom management seems to capture the extent to which teachers are able to promote student fidelity. Thus, in our study, it seems that teachers had to establish a stable classroom environment in order for learning to occur—or, more specifically, for students to benefit from the more specific components of the curriculum. Nonetheless, within the context of an effective learning environment, fidelity to content knowledge, modeling, and independence mattered for student success because students who were in classrooms where teachers implemented our curriculum well did twice as well on posttest measures as compared to students who were in classrooms where teachers did not use the curriculum with high fidelity. These data add more
credence to our findings and help to address the limitations caused by the absence of a comparison group and random assignment of teachers to conditions. That said, future studies with randomized assignment of teachers to condition and a more precise comparison group will help to further gauge the impact of our intervention over and above more traditional writing curriculum. Furthermore, future studies with cognitive apprenticeships might be designed to explore interactions among these variables, such as whether certain levels of teacher content knowledge and modeling are necessary to promote student independence.

Supporting Teachers in Their Use of a Curriculum

It seems intuitive that teachers’ pedagogical judgment makes a difference in their ability to foster higher levels of disciplinary writing and reading proficiency with students. In this study, we found that teachers who performed at the highest levels in adhering to our curriculum intervention also made successful adaptations—positive changes to the lessons that supported students’ independence with the strategies they were learning. To illustrate, teachers who implemented our curriculum with high fidelity were more likely to allow students to judge for themselves whether they needed more or less support from the teacher or peers, particularly in later investigations when the lessons called on students to work independently. One teacher encouraged students to move ahead at their own pace, and another invited students to work in pairs if they still needed help. Another example relates to decisions that teachers made with respect to changes in the timing or pacing of the curriculum in each of our six investigations. We found that teachers with high levels of fidelity tended to find ways to incorporate more content and made positive adjustments to the lessons, whereas adjustments made by teachers with low levels of fidelity sometimes detracted from instruction by losing instructional time to classroom management issues or digressions that did not address the learning outcomes for the lessons. Thus, our findings led us to make changes in our PD in subsequent years of our project, to help teachers distinguish between potentially positive and negative adaptations to our curriculum intervention. We share an in depth analysis of two reflective teachers who taught very different types of learners elsewhere (Monte-Sano, De La Paz, & Felton, in press).

CONCLUSIONS

The results of the current study suggest that a curriculum intervention focused on disciplinary strategies in reading and writing, in combination with teacher PD, can lead to substantial improvements in middle school students’ abilities to write historical arguments when controlling for students’ incoming abilities. These results are meaningful in that the schools in which we worked
included large numbers of culturally and academically diverse adolescent readers, which we contend indicates the broad potential for both a disciplinary literacy curriculum intervention and a discipline-specific cognitive apprenticeship model of instruction. By providing teachers and students with a scaffolded approach to working with primary historical sources, we have demonstrated that historical thinking is possible not only for college students and advanced high school students but also for students as young as in eighth grade and in classrooms with high numbers of struggling readers. We agree that writing argumentative essays in response to historical controversies does not in itself promote disciplinary thinking (Grant, Gradwell, & Cimbricz, 2004). However, argumentation is now recognized as one way for students to engage in literacies that are central to history (CCSSO, 2010). History teachers may engage in approaches that are similar to methods described here, knowing the benefits in teaching younger secondary students to engage in historical interpretation through literacy instruction.

ACKNOWLEDGMENTS

We are grateful to the teachers and students who made this research possible.

FUNDING

The research reported here was supported by the Institute of Education Sciences, U.S. Department of Education, through Grant R305A090153 to the University of Maryland. The opinions expressed are those of the authors and do not represent views of the Institute or the U.S. Department of Education.

NOTES

1 We use the term “intervention” throughout this article because disciplinary literacy was an area of need the cooperating school district and because the term captures the element of specificity that exists in our curriculum.

2 This percentage is in line with national data from the U.S. Department of Education, National Center for Educational Statistics (2012).

REFERENCES


**ABOUT THE AUTHORS**

**SUSAN DE LA PAZ** is an Associate Professor in the Department of Counseling Psychology, Higher Education and Special Education, and an Affiliate with the Department of Teaching, Learning, Policy and Leadership in the College of Education at the *University of Maryland*, College Park, MD 20742. She can be contacted at sdelapaz@umd.edu.
MARK FELTON is a Professor in the Department of Secondary Education, College of Education at San José State University, San José, CA 95192-0077. He can be contacted at mark.felton@sjsu.edu.

CHAUNCEY MONTE-SANO is an Associate Professor of Educational Studies in the School of Education at the University of Michigan, Ann Arbor, MI 48109-1259. She can be contacted at cmontesa@umich.edu.

ROBERT CRONINGER is an Associate Professor of Teaching, Learning, Policy and Leadership in the College of Education at the University of Maryland, College Park, MD 20742. He can be contacted at croninge@umd.edu.

CARA JACKSON is a Strategic Data Fellow and Assistant Director of Accountability at Urban Teacher Center, Baltimore, MD 21211. She can be contacted at caraj@urbanteachercenter.org.

JEEHYE SHIM DEOGRACIAS is an Accountability Analyst at Alexandria City Public Schools, Alexandria, VA 22311. She can be contacted at jeehye.deogracias@acps.k12.va.us.

BENJAMIN POLK HOFFMAN is a Doctoral Candidate in the Department of History at the University of Houston, Houston, TX 77204-3003. He can be contacted at bphoffman@uh.edu.
Head Note: Charlotte Forten was an African American from the North who went to the Sea Islands in Georgia at the end of the Civil War to teach newly freed slaves.

“Life on the Sea Islands,” The Atlantic Monthly

The first day at school was rather difficult. Most of my children were very small and restless. Some were too young to learn the alphabet. These little ones were brought to school because the older children had to take care of them and could not come without them. After some days, I found little difficulty in managing and quieting the tiniest and most restless spirits.

I never before saw children so eager to learn, although I had several years’ experience in Northern schools. Coming to school is a constant delight to them. They come here as other children go to play. The older ones work in the fields from early morning until 11:00 or 12:00, during the summer. Then they come into school, after their hard work in the hot sun, as bright and as excited to learn as ever.

The majority of students learn very quickly. Many of the grown people want to learn how to read. It is wonderful how a people who have been so long crushed to the earth can have so great a desire for knowledge, and such an ability for reaching it. One cannot believe that the arrogant English race, after centuries of such an experience as these people have had, would do any better. And I get angry at those in the North and the South who criticize the colored race as inferior while they themselves use every means in their power to crush and shame them, denying them every right and privilege, closing every path to improvement.

Daily the long-pressed people of these islands are demonstrating their capacity for improvement in learning and labor. What they have accomplished in one short year exceeds our greatest expectations.

By Charlotte Forten
May and June 1864

Head Note: Captain Hamilton, a Northerner in the U.S. Army, was posted in Florida after the Civil War. Here, Hamilton writes to the U.S. Army officer in charge of organizing and supporting soldiers about events at a school in the region where he is posted.

Florida, 1866

Dear Sir:

The night school has been frequently disturbed. One evening a mob called the teacher out of the schoolhouse. When the teacher showed himself, he was faced with four revolvers, and a man with a threatening expression of shooting him, if he did not promise to quit his job, and close the school.

The freedmen, who were now able to protect themselves, came to his aid promptly and the mob broke up.

About the 18th or 19th of the month, I was absent when an alarming riot took place at the school. The same mob threatened to destroy the school that night. The freedmen, learning this, gathered at the school in self-defense.

I understand that not less than forty colored men armed to protect themselves, but since their plans became known to the mob, the rowdy men only moved about in small groups, and were wise enough to avoid a collision.

[Hamilton’s signature]

Source: Excerpt adapted from a letter written by Captain C. M. Hamilton in 1866 to the Office of the Adjutant General in Washington, D. C.
Savannah, January 4, 1865

My Dear Sir:

I have been here for some days. The colored people did not seem to realize that they were free; this was not announced to them. The lack of supplies and the shortage of jobs—all these had the effect of causing our people to stand on the threshold of freedom like the rescued passenger of a ship on a barren sea-shore, wet and shivering with the cold blast of the storm. They needed encouragement, advice, and strength to go forward and act like free men.

I am happy to say their confidence has been growing every day and they seem to lose that fear which slavery had made a second nature. There are so many very intelligent colored persons in Savannah. We have been holding large meetings of the colored citizens. The interest has been great, and the promise of good being done is bright.

The Government has given us three large buildings:

1. “Bryant’s Negro Mart” (reads the sign over the door). It is a large three-story brick building. We are going to use it for a school.
2. The Stiles house on Farm Street, formerly used as a rebel hospital. We will also use it for school purposes.
3. A large three-story brick building on the lot next-door for a hospital for freedmen.

We have organized an Association called the Savannah Educational Association. Members are made up of the pastors and members of the colored churches. There are five very large colored churches in this city; four of them will seat one thousand persons each. Three have fine organs. That the colored people built such churches is astonishing. Hundreds of colored people are joining the Association.

We have talked to some of the most intelligent of the colored young men and women to find out their qualifications for teaching and selected nine. This makes use of the ability and intelligence held by the colored people, and gives them confidence and encouragement.

Refugees are continually coming in and filling up the city. Oh, how much books are needed! We could use right away a thousand spelling-books, if we had them.

James Lynch

Calhoun, Georgia, August 25, 1867

General:

We the Colored people of the town of Calhoun desire to call your attention to the Situation that now exists.

On the 16th day of the month, we held a political meeting of the Union Republican Party. The Colored people of the County attended it all together. Since that time we seem to have learned the hatred and spite of people who did not agree with the opinions that we shared in that meeting.

We see the hatred of these people in several ways. Their first act was to deprive us of the privilege to worship any longer in their Church. We have found a church of our own. But, they threaten us if we hold meetings in it.

There have been houses broken open, windows smashed and doors broken down in the dead of the night, men rushing in, cursing and swearing and firing their Pistols inside the house. Colored men have been knocked down and beaten for no reason and yet the police do not notice it at all. We would open a school here, but are almost afraid to do so since we have no protection for life or limb.

We wish to do right, obey the Laws and live in peace and quiet. But when we are attacked at the midnight hour, our lives threatened and the Laws fail to protect or help us, the only thing we can do is defend ourselves. Yet we wish to avoid all such conflicts.

We respectfully ask that a few soldiers be sent here. We believe it is the only way we can live in peace until the Election this fall.

[Twenty-four signatures]

Source: Excerpt adapted from a letter written by 24 African Americans from Calhoun, Georgia, to the Commander of the Third Military District. August 25, 1867.
APPENDIX B

Observation Protocol for First Day of Investigation 1

<table>
<thead>
<tr>
<th>No.</th>
<th>Questions</th>
<th>0</th>
<th>1</th>
<th>2</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>At the beginning of the lesson, does the teacher present history as investigation and students as detectives looking for clues?</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Does the teacher organize students to work in partners or in groups (to support each other in reading) when they look at documents? 0 = No; 1 = Somewhat but may be ineffective or inconsistent; 2 = Yes, sufficient</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Comment:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Does the teacher use Handout A.1 to guide the investigation?</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>Does the teacher give students Document 1 to analyze (Handout A.2)? 0 = No; 1 = Partial use; 2 = Yes, sufficient</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>Does the teacher give students Document 2 to analyze (Handout A.3)? 0 = No; 1 = Partial use; 2 = Yes, sufficient</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>Does the teacher ask students to write their initial hypothesis, document analysis questions, and revised hypotheses? 0 = No; 1 = Yes, at least one step; 2 = Yes, each step If some, but not all, which parts of the handout did teachers ask students to complete?</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>Does the teacher refer back to earlier hypotheses when students revise their hypotheses? 0 = No; 1 = Yes, one time; 2 = Yes, both times</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>Does the teacher ask students what evidence supports their revised hypotheses? 0 = No; 1 = Yes, one time; 2 = Yes, both times</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>Does the teacher ask students which pieces of evidence were more convincing in their investigation (or which were the best clues)? 0 = No; 1 = Yes, asks one student; 2 = Yes, asks more than one student</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>Is the discussion of these ideas (of which evidence was most convincing) thorough? 0 = No; 1 = Somewhat (teacher responds to at least one student’s idea); 2 = Yes (teacher and/or students respond to more than one student’s ideas)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
### Appendix B. (Continued)

<table>
<thead>
<tr>
<th>No.</th>
<th>Questions</th>
<th>0 No</th>
<th>1</th>
<th>2 Yes</th>
</tr>
</thead>
<tbody>
<tr>
<td>11</td>
<td>Does the teacher present history as interpretation at the end of the lesson (Handout A.4)?</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0 = No; 1 = Partial attention or inconsistent or inaccurate; 2 = Yes, sufficient</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Comments:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>Did the teacher implement the lesson plan for this day without making changes in any way?</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>If so, how and when did the teacher alter the plan?</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Comments:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>Do you know why the teacher altered the plan?</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Comments:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>Did the teacher seem to understand the ideas and strategies in this lesson sequence?</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Comments (How do you know? What do you base your assessment on?):</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>Did the teacher work effectively with students to help students understand the ideas and strategies in this lesson sequence?</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0 = No; 1 = Somewhat; 2 = Yes</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Comments (How do you know? What do you based your assessment on?):</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>Could the students complete the tasks without help from researchers?</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0 = Few to none of the students on few of the tasks; 1 = Yes, some of the students, some of the time; 2 = Yes, most of the students, most of the time</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Comments (What kinds of help did students need?):</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>Did the teacher teach the lesson without needing any help from researchers?</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0 = No; 1 = Yes, some of the time; 2 = Yes, most of the time</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Comments (What kinds of help did the teacher need?):</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**TOTALS**

**GRAND TOTAL** %

### Notes

Construct 1—Building understanding of strategies; see questions 15 and 16 on this day of the lesson.

Construct 2—Promoting independence in using strategies; see questions 1 and 8–11.

Construct 3—Building historical and disciplinary knowledge; see questions 3–7, 12–14, 17, and 18.

Construct 4—Classroom management; see question 2.
We asked observers to create a post-observation summary in one paragraph and to sum up the highlights of what he/she saw with regards to . . .

- teaching historical thinking and annotation,
- working productively with students,
- the adaptations teachers made to the lesson plan,
- the kinds of help students and teacher needed, and
- use of technology.

APPENDIX C

Sample Pretest and Posttest Plan and Essay—Below-Basic Reader (Color Figure Available Online)

I think that African Americans were not free after the Civil War. Because in Document Two, it says that they had "houses broken open, windows smashed and doors broken down in the dead of the night." Also, it says "men rushing in, cursing and swearing and firing their pistols in side the house." "Colored men have been knocked down and beaten for no reason." Even if they have been freed, other people aren't being fair to the colored race. An example of unfairness was that "Their first act was to deprive them of the privilege to worship any longer in their church." Another example is "When they were attacked at the midnight hour, their lives threatened and the laws said to protect or help them in time of need. So I think that African Americans were not free after the Civil War.

Length = 146 words, Evidence = 3, Perspective = 0, Context = 1, Holistic = 1 at pretest
I am comparing the opinions of Charlotte Foster and Captain Hamilton on if African Americans were free after the Civil War. And the two documents took place around the 1860s in Georgia and Florida.

Charlotte Foster thinks that the African Americans were not free after the Civil War. I know this because in the first document, it says, “And I got angry at those in the north and the south who criticized the colored race as inferior while they themselves were oppressors in their power to crush and degrade them denying them every right and priviledged, closing every path to improvement which I mean your rights and privileges.”

I think that they were not completely free because some rights and privileges were taken away from African Americans and they were not treated equally.

Charlotte Foster thinks that the African Americans were not free after the Civil War. I know this because in the first document, it says, “And I got angry at those in the north and the south who criticized the colored race as inferior while they themselves were oppressors in their power to crush and degrade them denying them every right and privilege, closing every path to improvement which I mean your rights and privileges.”

In contrast, Captain Hamilton thinks that African Americans were free after the Civil War because in the second document, he says, “Some free negroes, who were now able to defend themselves, came north and promptly the law broke up. This meant that they had weapons and so they could not be a slave and they could now use a weapon so they were free.”
APPENDIX D

Sample Pretest and Posttest Essay—Advanced Reader (Color Figure Available Online)

Even though African Americans were free from slavery and working without pay, they were not free from oppression and racial injustice. I have reason to believe that they were not free because they had a different color skin and people still saw them as slaves. Even though they were free, physically, they were not free emotionally. The north did not want slavery but the south did so that is why the south created laws that stopped them from voting, eating with other people, drinking from the same water fountain, and riding on the bus were they wanted to set.

Length = 101 words, Evidence = 0, Perspective = 0, Context = 1, Holistic = 2 at pretest
In 1861, Confederate states burned down Fort Sumter in South Carolina. This marked the beginning of the Civil War. The North was against the South for many reasons. The main one was because of slavery. In Document 1, the author was shocked that many colored were unaware they were free. So the author helped them by creating an group called the Savannah Educational Association. They built churches, schools, and hospitals. This was all possible by the Emancipation Proclamation. In Document 2, the author is unhappy about how blacks in California are being treated. He says there noses are being broken into, windows smashed, colored men being beaten for no reason at all. In this document the author asks for soldiers be sent to California for peace. My point of view on this matter is I believe that African Americans were not free. Sure they weren't slaves anymore, but they were restricted from doing certain things. Things like voting. Also, the southern governments would use a variety of techniques to stop them from voting, like poll tax, grandfather clause, literacy test, and segregation. I think African Americans after and before the civil war were not free.