Received Date : 29-Sep-2015
Revised Date : 25-Feb-2016
Accepted Date : 02-Mar-2016
Article type : Article

A Historical Writing Apprenticeship for Adolescents: Integrating Disciplinary Learning With Cognitive Strategies

Susan De La Paz

University of Maryland, College Park, USA

Chauncey Monte-Sano

University of Michigan, Ann Arbor, USA

Mark Felton

San Jose State University, California, USA

Robert Croninger

University of Maryland, College Park, USA

Cara Jackson

Urban Teachers, Baltimore, Maryland, USA

Kelly Worland Piantedosi

District of Columbia Public Schools, Washington, DC, USA

ABSTRACT

This study explored the extent to which an 18-day history and writing curriculum intervention, taught over the course of one year, helped culturally and academically diverse adolescents achieve important disciplinary literacy learning in history. Teachers used a cognitive apprenticeship form of instruction for the integration of historical reading and writing strategies and content learning with the goal of improving students' historical argument writing. The intervention had positive and significant results for each writing outcome. After controlling for variables associated with students' incoming abilities, the researchers found

moderate to large effects for all participants. Relative to basic readers in the control condition, those participating in the intervention scored higher in historical writing and writing quality and wrote longer essays; these results translate into effect sizes of .45 on basic readers' historical writing, .32 on their overall writing quality, and .60 on the length of their papers. Teachers implemented the reading and writing curriculum intervention with high levels of implementation fidelity, leading the researchers to explore additional factors that contributed to students' success after accounting for teacher effectiveness. The results indicate further benefits dependent on the degree to which students completed the curriculum.

Over the past 10 years, there has been growing recognition that by adolescence, writing to learn and learning to write in school must connect learners to ways of knowing in the disciplines (Carter, Ferzli, & Wiebe, 2007; Moje, 2008). Although some debate the most appropriate aim for secondary content area literacy instruction (Conley, 2012; Draper, 2008; Shanahan & Shanahan, 2012), with calls for writing to support knowledge acquisition (Heller, 2010), there appears to be growing consensus that as students progress through the curriculum, they should write not only to demonstrate content area learning but also to grapple with domain-dependent and intellectually challenging issues (Bain, 2012; Beaufort, 2004; Moje, 2008; Stevens, Wineburg, Herrenkohl, & Bell, 2005). Secondary literacy and content area learning have become inextricably interlinked, with academic progress increasingly dependent on the acquisition of specialized knowledge and skills and distinct purposes in literate, scientific, and historical communities (Greenleaf, Schoenbach, Cziko, & Mueller, 2001; Torgesen et al., 2007).

Recent standards initiatives have added urgency to teaching writing in content area classrooms. The emphasis on writing argument across content areas in the Common Core State Standards for English Language Arts (National Governors Association Center for Best Practices & Council of Chief State School Officers, 2010) and the prominence of inquiry, disciplinary thinking, and communicating conclusions in the College, Career and Civic Life (C3) Framework for Social Studies State Standards (National Council for the Social Studies, 2013) encourage history teachers, in particular, to develop students' disciplinary thinking and writing. No longer is literacy development the official purview of English teachers only, nor is literacy simply a matter of developing facility with general reading and writing practices regardless of content. Instead, literacy is now framed as a crucial feature in any effort to help students understand and develop knowledge in the disciplines. Because disciplinary literacy includes reading and writing as well as conceptual understanding and procedural knowledge of a discipline, the literacies

demanded are not generalizable; they vary by discipline. Rather than posit an absolute advantage for domain-specific literacy alone, we argue that discipline-specific literacy strategies are best learned in tandem with domain-general literacy strategies to help young or struggling adolescent learners write coherent, persuasive texts that cite, interpret, and argue from evidence according to sound disciplinary standards and practices.

Defining Historical Writing

In this study, we integrate content and literacy by focusing on reading, thinking, and writing practices in the context of history, with improved writing as the ultimate goal. We refer to our central outcome of interest as *historical writing* and define it as an interpretation based on evidence that makes an argument about another place and time. These interpretations strive for understanding of the past, often by making arguments about cause and effect (e.g., Coffin, 2006) or change and continuity (e.g., Seixas, 2006). Writing is a visible representation of historians' thinking and the process of developing claims based on analysis of the historical record.

The public display of evidence (via footnotes) and where it comes from enables historians to substantiate their arguments. Historical writing is rooted in evidence that takes many forms—diary entries, tax records, speeches, paintings, photographs, objects, and so forth—but the historical record is incomplete. We do not have all records from every perspective at any given point in time; therefore, historians do some amount of imagining and make tentative conjectures based on these historical sources, or traces of the past (e.g., Hexter, 1971).

Reading is integral to historical writing because historians engage in detective work to understand the meaning of the evidence they use to develop the interpretations they share in writing. This largely involves moving beyond what is literally stated in a text to uncover the subtext of each source through questioning. Because historical sources were created in another time and place, historians must reconstruct the circumstances of their creation. Wineburg's (1991) seminal work uncovered particular aspects of disciplinary thinking that goes into the analysis of evidence: sourcing, contextualizing, and corroborating. He found that historians look for clues about the motivations and experiences of the authors who created the sources that historians analyze and the degree to which authors are reliable for the inquiry at hand (sourcing). Likewise, historians consider what was happening at the time and place in which the author created the source to situate historical sources in their context (contextualization). Historians compare and contrast sources to determine what conclusions they can reliably make about the

questions they pursue (corroboration). Also, historians address counterevidence and different perspectives rather than cherry-picking evidence that supports their claim, a process that often leads to altering the claim to reflect the evidentiary base (Hexter, 1971). These ways of reading and thinking are apparent in students' historical writing as well (Monte-Sano, 2010).

Because the arguments that historians convey in writing are grounded in the process of interpreting evidence, historical writing necessarily embeds disciplinary thinking and reading (Monte-Sano, 2011; Young & Leinhardt, 1998). The ultimate goal is to convey an evidence-based interpretation, or argument, in writing. In the process of writing, therefore, historians ask questions, read and analyze primary (and secondary) sources, critique and weigh evidence, consider multiple perspectives, sort and organize ideas and evidence, and construct evidence-based claims (Nokes, 2013).

Benefits for Students

History classes are prime sites for teaching argument writing given the centrality of evidence-based interpretation to the discipline. In-depth investigation of historical events and people also provides the opportunity for students to understand a topic and remember details about it (e.g., Reisman, 2012). Historical writing generally orients students toward history as an interpretive discipline grounded in analysis of evidence, rather than one focused on factual recall. When taught to write their own interpretations, students are given a window into the discourse that emerged concerning historical events, they develop a sharper sense of the multiple and sometimes conflicting perspectives on interpreting the meaning and significance of these historical events, and they learn to appreciate that historical knowledge is constructed rather than received or uncovered (Monte-Sano, 2008; VanSledright, 2002).

Research also tells us that writing essays in history can improve students' mastery and understanding of factual information (Bangert-Drowns, Hurley, & Wilkinson, 2004; Smith & Niemi, 2001). We know from prior studies that writing essays in history can enhance students' ability to integrate content from sources with their own thinking (Wiley & Voss, 1999; Young & Leinhardt, 1998), and promote historical thinking (Monte-Sano, 2010). There seems to be a connection between writing arguments and greater attention to source information (Le Bigot & Rouet, 2007; Stahl, Hynd, Britton, McNish, & Bosquet, 1996; Wiley & Voss, 1999). Disciplinary writing thus appears to be a promising approach to improving secondary students' writing and understanding of history (e.g., De La Paz, 2005; Moje, 2008).

Challenges

Despite the natural fit between writing and history given the discipline's structure and purpose, work on writing and interpretation have not been commonplace in school. Increasingly, history classrooms have embraced primary source–based investigation and inquiry (e.g., Ragland, 2007); however, writing is not a regular part of students' social studies experience in school. Only 32% of eighth graders attested to writing long answers to questions or assignments for history/social studies (National Center for Education Statistics, 2015) at least weekly on a recent National Assessment of Educational Progress survey; in contrast, 64% wrote short answers to questions on a weekly basis. When teachers assign reading and writing in secondary history classrooms, the focus typically involves reading comprehension and summary of information (Kiuhara, Graham, & Hawken, 2009), as well as the use of textbooks as authoritative sources of information (Bain, 2006). Nokes's (2010) observational study of eight high school history teachers provided further support for these challenges: Teachers allocated very little class time to using primary sources. When they did so, they typically read the sources to their classes and explained what the sources meant in relation to lecture material, rather than allowing students to analyze their meaning.

Such approaches to history instruction do not give students the opportunity to analyze, question, and weigh artifacts from the past nor to construct their own interpretations of historical events and people. The presentation of history as static information encourages students to see the subject as a given set of fixed stories and relegates them to passive reception; such an epistemic stance leaves no room for analysis or interpretation and inhibits students' historical writing (Monte-Sano, 2008). Finally, history teachers are not typically prepared to teach writing (e.g., Ragland, 2007), nor are materials that support this kind of disciplinary thinking widely available; instead, textbook-based instruction dominates (Bain, 2006).

Instruction That Works

Studies of high school U.S. history classrooms have identified factors to improve historical writing: investigative questions that present history as an inquiry-oriented subject and call for argument, reading contrasting historical sources with support for comprehension and historical thinking (e.g., reading questions or annotation prompts), giving students opportunities to construct interpretations and support them with evidence (including class discussion), and combining explicit instruction with guided practice and feedback (Monte-Sano, 2008; Wiley &

Voss, 1999; Young & Leinhardt, 1998). Larger studies have applied these concepts via explicit instruction in historical thinking and persuasive writing as students worked with primary sources in middle school (De La Paz, 2005) and high school classrooms (De La Paz & Felton, 2010). Together, these studies illustrate that developing students' historical writing requires attention to the disciplinary nature of reading, writing, *and* explicit instruction. Studies that take a cognitive apprenticeship approach (Brown, Collins, & Duguid, 1989) have helped students learn general persuasive writing (De La Paz, 2005) or historical writing (De La Paz et al., 2014), but not both simultaneously. In addition, these studies of a cognitive apprenticeship approach to teaching historical writing have been successful when working in a handful of schools only.

Our Purpose

In this study, we sought to determine whether a cognitive apprenticeship approach to history instruction could be used to support growth in students' general writing and historical writing at a large number of schools. Thus, in our current study, we continued to emphasize historical writing but also focused on general argument writing skills to determine whether students could grow in both general and disciplinary literacy in one intervention. In earlier work reporting on year 1 (De La Paz et al., 2014), results indicated students' need for support in basic argument writing skills and historical writing, and the current study shares our final curriculum intervention and professional development (PD) from years 2 and 3 of the project. In the current study, we tested whether combining instruction in general argument writing with instruction in historical writing promoted more coherent and comprehensive argumentative writing in history. In effect, this study highlights the need to support basic literacy alongside disciplinary literacy when addressing literacy in domain-specific settings.

Research Questions

In the current study, we asked three questions:

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

2. Do students with advanced, proficient, and basic reading proficiency levels all benefit from the yearlong instruction?

3. How does fidelity of implementation with the core components of cognitive apprenticeship relate to student learning?

Method

Design

This study used a quasi-experimental design, comparing student writing from teachers who volunteered to participate in our PD and use our curriculum intervention against student writing from control teachers who administered pretests and posttests and used the cooperating school district's pacing guide and lesson materials for instruction. We were unable to randomly assign teachers or students to conditions due to requirements set by our funding. Full details about our PD are available elsewhere (Monte-Sano, De La Paz, & Felton, 2014a); however, in the current study, we met with teachers for 66 hours of focused PD across 11 daylong sessions in one year, and in the next, we met with teachers for 60 hours across 10 sessions.

Setting and Participants

We worked with a large school district on the border of a major city with urban, suburban, and rural communities in the mid-Atlantic region of the United States. The district serves socially and ethnically diverse students: 45% of the students receive free or reduced-price meals, 8.5% receive instruction in English for speakers of other languages, and the majority of the students are black or Hispanic/Latino/a. Each year we worked with different eighth-grade U.S. history teachers and their students, choosing schools where the district had identified 15–30% of the student population as significantly below grade level in reading. Although these schools had significant numbers of struggling readers, most students were proficient or advanced readers.

We worked with 19 teachers and 2,143 students in eight schools one year and 17 teachers and 2,151 students in 11 schools the next year, although the final number of eligible participants was slightly lower due to absences from school during our testing. Our funding was for the development of a curriculum intervention that could produce pilot data on the potential benefit for our approach; therefore, we recruited teacher volunteers at target schools for participation in our project. Teacher participants had a range of experience (e.g., some were new to teaching or to teaching social studies) and experience in teaching at the middle school level. Control teachers also had varying types of experience; some chose not to participate in our project because of administrative responsibilities or other commitments (e.g., participation in other PD projects, coaching for intramural sports) or because they believed they were already capable of teaching

their students to read and write from historical documents, which were included in the district pacing guide and the focus of a related district initiative on disciplinary literacy.

Our curriculum intervention was taught over one academic year, and we evaluated outcomes by comparing pre and post essays with those written by a business-as-usual control group. Because we could not analyze all of our data, we followed a stratified random sampling plan to select a representative sample from teachers in both conditions, choosing about the same overall number of students from each teacher while simultaneously balancing gender, ethnicity, and the level of students' incoming reading and writing abilities. Thus, this report is based on data from 36 teachers and 1,029 students during two years of our project (see Table 1 for participants' characteristics). A total of 22 teachers participated in our treatment condition (working with 645 students), and 14 teachers participated in a comparison condition (working with 384 students). Social studies teachers were expected to adhere to a pacing guide that laid out specific information in U.S. history to cover. They also administered multiple-choice exams at the end of each semester as required by the district. Most teachers were accustomed to textbook-based instruction that emphasized factual recall; therefore, this curriculum intervention posed a major shift in social studies instruction.

[COMP: Please insert Table 1.]

We computed *t*-tests and chi-square analyses for all student-level variables to examine equivalence between the intervention and control groups. This was conducted to account for any potential differences between the two groups on background characteristics. We found two significant differences (at $p < .05$) between the intervention and control groups on demographic characteristics: The control group had a greater proportion of Asian students (7.0% vs. 4.2% in the intervention group) and a greater proportion of English learners (7.3% vs. 4.0% in the intervention group). We did not find significant differences in the proportion of the samples who were white, Hispanic, African American, Native American, or of other race. There were no significant differences between the intervention and control groups on preintervention state reading assessment proficiency levels or the proportion of students with Individualized Education Plans (IEPs).

Curriculum Intervention

We set out to bridge the gap between teaching literacy and teaching history by constructing a curriculum that integrates the two by focusing on writing, argument, and thinking practices in the

context of history. We adopted a cognitive apprenticeship approach to instruction (Brown et al., 1989) with both teachers and students because writing historical arguments requires the ability to coordinate multiple cognitive processes when thinking about historical content: conceptual knowledge of history; disciplinary acts when analyzing evidence (e.g., sourcing, contextualizing, corroborating; Wineburg, 1991); and topical, factual information.

We based our cognitive apprenticeship model on principles of strategy instruction (cf. self-regulated strategy development; Harris & Graham, 1996) when teaching students to access and evaluate historical content while reading and to engage in argumentative writing through a series of carefully designed scaffolds for reading and writing. During the first half of the year, teachers described foundational concepts about historical reading and writing and modeled how to use heuristics, using them in a way that was visible to students and by thinking aloud during modeling. In the process, teachers worked with students to articulate historical reading, thinking, and writing practices in the context of each investigation.

The rest of the year, teachers primarily focused on students' application of the strategies, with an increasing focus on how they were to manage the reading and writing processes on their own, through collaborative and independent practice stages of instruction. Although we did not employ all elements of self-regulated strategy development (e.g., self-regulation did not include guiding students with regulatory self-statements), teachers frequently asked students to reflect on the historical concepts and practices and how underlying components of the intervention related to and supported the overall goal of writing argumentative essays, and asked students to set personal goals.

Rather than addressing skills as discrete or decontextualized, we sought to maintain the complexity of historical writing by situating students' learning in the context of historical inquiry (e.g., working with conflicting primary sources to investigate a central question) that required their participation in reading, thinking, and writing activities in an integrated, authentic way (cf. Brown et al., 1989). So students could gain access to these practices, teachers initially modeled historical ways of reading, thinking, and writing in situ—as students participated in historical inquiry—by making their thinking explicit for and visible to students as they used those disciplinary practices (Collins, Brown, Holum, 1991). During PD, we modeled and discussed the differences between simply telling students what to do and actually performing the practice and externalizing the thinking that goes into using the practice, to allow students to grasp what is

necessary to use the practice successfully in a way that directions alone cannot. After we modeled each practice, teachers rehearsed modeling each practice and planned how to adapt instruction for different types of learners.

The social studies curriculum in the cooperating school district focused on U.S. history in grade 8, with content standards that began with a focus on the Revolution and ended with Reconstruction. We worked with this content and grade because the district felt it was the area most in need of improvement within social studies. We chose six topics in collaboration with district personnel and created a three-day lesson sequence for each (see Table 2). District negotiations led to an agreement that teachers and students would be available for 18 days of instruction, which we believed was the minimum required for students to master key disciplinary literacy practices (e.g., reading, discussing, and evaluating evidence from sources; planning and writing) within the cognitive apprenticeship, based on our earlier work (De La Paz, 2005). We referred to each lesson sequence as an investigation, framing the work of history as inquiry, and each investigation began with a central, controversial historical question that served as the driving purpose of students' work.

[COMP: Please insert Table 2.]

Within each three-day investigation, students learned and used key historical reading, thinking, and writing practices with the help of scaffolds designed to articulate and reinforce these practices. Day 1 of the first three investigations involved explicit instruction in reading and annotating the documents, with a particular focus on the historical background for each controversy and basic comprehension of the sources. On day 2, students read and analyzed documents and learned how to think historically about sources, considering the influence of author, context, and authors' facts and examples. Day 3 involved planning and composing an essay using a visual illustration of the underlying text structure for a five-paragraph essay, two sample essays with opposing arguments and the exemplary text structure, and an extended set of sample phrases and sentence starters (e.g., "After reading information from both sides...," "His/Her quote supports my argument because…") for introducing ideas and quotations when writing. The sample phrases also made visible how students could engage in sourcing (e.g., "This author is a better source of information because…") and how to judge evidence (e.g., "Another event in history that relates to this was when..."). Over the course of the year, the focus of the

middle day shifted as students gained facility in discussing and evaluating evidence and moved toward planning and composing their essays with a reduction in teacher assistance.

Stages of Instruction

In keeping with a cognitive apprenticeship approach to instruction, teachers prepared students to learn by developing their background knowledge for disciplinary thinking processes using the cognitive scaffolds for reading primary sources, and planning historical essays in the first three investigations. Teachers initially modeled how to use these supports, supported students through guided practice in using them, and then promoted their independence by providing additional, more challenging forms of practice, ultimately fading the supports that were built into the curriculum. Most students worked without overtly using either scaffold, reading, planning, and writing independently in two days during the last investigation.

In sharing our curriculum, we now more fully describe the disciplinary scaffolds and explain how they were used in our lessons to help students learn strategies for historical reading, analysis, and argumentative writing: The mnemonic IREAD (defined in the next section) was used to guide students' reading and annotations, and the "how to write your essay" (H2W) text structure and sample essays were used to guide students' planning and composing. We recognize that there are risks in boiling complex processes down into concrete tools for students' use; for example, students and teachers may learn to follow discrete steps without gaining foundational understanding (Westhoff, 2009). In our curriculum, teachers modeled use of each scaffold flexibly, emphasizing that the overarching purpose of instruction was to engage in disciplinary reading and writing while they guided students' attempts at using each heuristic independently. Highly structured learning opportunities such as these have been effective in research with academically diverse students (De La Paz, 2005). We associate each scaffold with core teaching practices in history education that are necessary for teaching historical writing. Yet, for most teachers, these forms of support and the practices associated with them represent a departure from conventional social studies instruction (Cuban, 1991).

Overview of IREAD

We took lessons from Wineburg's (1991) research to heart and initially constructed IREAD to focus on the subtext of historical texts, emphasizing inferences about the texts rather than their literal meaning. Yet, we found that students and teachers sometimes avoided reading the entire text because doing so was challenging. Instead, they searched for specific clues around the

periphery of the text, such as the author and date of creation, without focusing on the body of the text. Even if they read the text, students often had little to no basic understanding of it, which limited their ability to understand the subtext and draw inferences, even if they noticed features such as author and date. We realized that we needed to balance reading comprehension and historical reading strategies more evenly for the students to be successful (for a poster version of this scaffold, see Appendix A, which is available as supporting information for the online version of this article).

To support generic comprehension, we used *IR* ("Identify the author's purpose" and "Read each paragraph and ask about the author's main ideas") to prompt students to identify and summarize what the author wrote (e.g., Jitendra & Gajria, 2011). One cue for *I* is to consider what the author would say in response to the historical question. Certainly, the authors of historical texts did not write with the questions we ask in mind; however, this cue supports students' reading comprehension by having them connect the text to the investigative question, providing guidance and purpose as students read.

The next two parts of the mnemonic, *EA* ("Evaluate the author's reliability" and "Assess the influence of context"), prompt students to source and contextualize texts (Wineburg, 1991) so they might begin to regard historical texts as the product of an author with intentions and as situated in a different time and place. We added the final prompt, *D* ("Determine the quality of the author's facts and examples"), to highlight the idea that authors use evidence to support their own arguments. Although not all primary sources are arguments, we selected ones that are so students have an opportunity to read arguments that model the kind of writing they are asked to do.

Together these prompts encouraged students to analyze and critique the texts rather than amass information about them. IREAD embeds a process of annotation to help struggling students notice specific aspects of texts and track their thinking. In previous work, we observed the annotation process used effectively with advanced students (Monte-Sano, 2010). We prepared IREAD in a foldable version for students, in which the front flap included questions to prompt students' historical thinking and the inside flap directed students to make specific notations for each step of IREAD or way of thinking (e.g., underline anything that has to do with the setting or context; Monte-Sano, De La Paz, & Felton, 2014b). Teachers called students' attention to using this tool flexibly to support the overarching purpose in reading critically in the

penultimate investigation, and students used their own foldable in each investigation until they were prompted to recall the meaning of IREAD and engage in the underlying processes without a physical reminder in the sixth investigation.

Overview of H2W

We created a graphic display of a particular form of argumentative text visually (see Appendix B, which is available as supporting information for the online version of this article), based in part on prior research by Meyer, Brandt, and Bluth (1980), who found that text structure allows readers to identify and remember top-level or central information from text. Explicit signals (e.g., topic statements, summary statements, keywords) cue text structure and the location of expository content. Good readers who detect these cues can remember more ideas when reading expository texts than readers who do not search for or identify text structure (Meyer & Freedle, 1984). Moreover, intervention work on the use of text structure (beginning with Englert, Raphael, Anderson, Stevens, & Fear, 1991) has indicated that providing students with direct instruction on how expository ideas are organized into text structures is a successful scaffold for helping students write better essays.

We extended this work by creating a disciplinary text structure, representing it in a graphic organizer entitled "How to Write Your Essay" (in the H2W text structure), which included essential components of historical arguments and information signaling how to organize essential components in the composition. We embedded a list of transition words and phrases (e.g., "this point makes sense"; "when all of the facts on both sides are considered") that were shown in relevant categories for historical writing (e.g., evaluating a quote, wrapping things up) based on the success of this type of scaffold in our prior work (De La Paz, 2005). The H2W graphic organizer reminded students to use evidence that they had identified and evaluated through reading and discussion, when writing their historical arguments.

Because most students did not have experience in writing historical arguments before this study began, we provided two sample essays to clarify what each aspect of an argumentative essay actually looks like. Teachers reviewed each major element in the H2W text structure (e.g., students could begin rebuttal paragraphs by choosing "the strongest reason, quote, or other evidence that goes against your argument but explains the other perspective") and helped students identify corresponding textual examples in each essay. The essays were written from opposing perspectives but with the same text structure, and both provided students with

examples of clear, evidence-based arguments. The sample essays were actual compilations of partial essays that were written by previous students, to make them more realistic (as opposed to being too advanced) for students to grasp.

To illustrate the complexity of historical inquiry, midway through the year for investigation 4, we asked students, "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?" In this investigation, students read a letter and a pamphlet written by Cherokee leaders on opposite sides of the debate, sources that demonstrated the complexity of the historical debate. We purposefully avoided using one Cherokee source and one from a U.S. government official to prevent students from automatically taking one side or the other. Instead, we selected sources that demonstrate the complexity of historical debates (i.e., that there often is no clear-cut, black-and-white answer). This example showed students that the Cherokee were not one united, homogeneous group and that the dilemma was which path would have allowed the Cherokee Nation to thrive, given what else was happening at the time.

More than previous lessons, this investigation required students to connect their reading, thinking, and writing to the study of history. By this time, they had learned the major strategies for reading and writing, and they now had a chance to use them together. As students went from developing background knowledge to reading and historical thinking, and then to planning and composing, they could see that these activities were related, with each process contributing to the final goal of writing evidence-based arguments in response to a historical question. Reading and historical thinking guided students toward an interpretation that was best supported by the evidence, which became the basis for their written argument.

As students annotated the Cherokee letter and Elias Boudinot's pamphlet, they also engaged in prewriting. When students planned their essay, they were prompted to reread primary sources and reconsider the evidence in light of the question—in other words, to read critically. In thinking about the central question and practicing these literacy strategies, teachers and students discussed that each strategy was not an end in itself but instead part of a larger thinking process that leads to writing an evidence-based argument. In terms of disciplinary literacy, this investigation emphasized evaluating evidence rather than accepting texts at face value, along with the practice of planning and composing a full essay.

It is important to note that teachers helped establish background knowledge for each investigation (although they did not do this at either pre- or posttest) with timelines, maps, video clips, and historical facts about the time period and events that related to each controversy. Teachers also reviewed up to five vocabulary words (e.g., *abolitionism*, *abolitionist*, *oppressor*, and *persecution* in investigation 5) before students read and annotated the primary sources. Teachers often reiterated the goal when reading was to identify and evaluate evidence in each document, in preparation for responding to the historical question. As students gained mastery in disciplinary reading and writing, teachers shifted from guiding students in step-by-step actions to reminding them of the supports, and suggesting time limits for most to follow as they worked independently. Teachers circulated among students as they worked, answering questions and offering support to struggling students (e.g., asking students to explain evidence that they planned to use and how it supported their argument, encouraging them to write fewer paragraphs if they were spending too much time on any one paragraph). We asked teachers to save time for reflection, even if some students had not finished writing their essays. We felt that it was valuable for students to see that others could interpret the same issue differently, to celebrate successful student writing excerpts, and for each student to determine goals for his or her future writing. Finally, during PD, teachers analyzed four or five students' work over the year, targeting different types of learners, to note strengths and areas of improvement in student writing and to set writing goals.

PD

The current study addresses challenges related to curriculum implementation, because in our prior work, half of the participating teachers could not reliably implement our curriculum intervention (De La Paz et al., 2014). Extensive teacher PD seemed necessary because of multiple goals in our program, such as the use of historical inquiry as a platform for learning disciplinary thinking and content, and the use of cognitive apprenticeship to teach students to independently engage in reasoning and writing strategies, which contrasted to many teachers' expectations for both the content and the focus of their instruction. Therefore, in this follow-up study, we redesigned our PD to first develop a shared view on developing what disciplinary reading and writing in history might mean for adolescent learners, in a learner-centered environment. We built on this foundation with lessons and materials from our curriculum intervention, employing many features of cognitive apprenticeship (modeling, practice,

feedback, and reflection) to promote teacher understanding and independence with key teaching strategies.

We used Grossman, Hammerness, and McDonald's (2009) framework of sharing representations, decompositions, and approximations of practice to demonstrate how to use the specific scaffolds and the approach to instruction in our intervention for each investigation. Using this guide, in each PD session, we modeled the use of investigation materials, debriefed the key elements and talked through how teachers might enact these elements, and gave teachers opportunities to practice teaching key aspects of the investigation to their peers. Practice sessions involved teachers working in small groups of four or five, taking turns modeling (including thinking aloud while using the strategy) and coaching with the materials, sharing feedback, and brainstorming how to use the curriculum effectively in their classrooms. In this way, PD sessions included modeling the use of disciplinary literacy strategies as well as practice in using these types of support so teachers could learn a cognitive apprenticeship approach to instruction in the context of historical inquiry. Finally, after they began using the curriculum, teachers analyzed students' written work and reflected on what students were learning, to consider how to respond to challenges they were seeing and their role in teaching the lessons. We believed this PD would enhance teachers' subject matter knowledge, provide extended learning time, actively engage teachers, and link well with what they were asked to do (Wilson, 2009).

Data Sources

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 pre- and posttest. Students had not learned about the post–Civil War era before either test. We created two forms of this test to allow us to counterbalance the measures (De La Paz et al., 2014). Both forms asked the same question, but each used a different document set. Form A consisted of two adapted letters, one from an 1864 issue of *The Atlantic Monthly*, "Life on the Sea Islands" by Charlotte Forten, and the other from Captain C.M. Hamilton to the Office of the Adjutant General in Washington, DC in 1866. Both letters describe events and perspectives related to schooling for African Americans after the Civil War. Form B consisted of two documents that provided students with information about African Americans' lives and opportunities to pursue individual freedom (see De La Paz et al., 2014). 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 some students were randomly assigned to respond to Form A at pretest and others to Form B at pretest, both within condition and within teachers. 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. We also computed analyses of variance on each of our dependent measures after the study ended and found no significant effects for test form.

Researchers in history education have used similar tasks to assess students' historical thinking and writing (cf. Rouet, Britt, Mason, & Perfetti, 1996; Seixas, 2006; Young & Leinhardt, 1998). Our approach is consistent with notions regarding analysis of evidence, use of evidence to construct interpretations of the past, and communication of arguments in writing. Although 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 following guidelines by Wineburg and Martin (2009). We excerpted the sources, focusing on segments that were most relevant to the question, so each source document was no more than one page. We created a headnote 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 it. Finally, we substituted simpler vocabulary where necessary to attain Lexile scores appropriate for sixth graders because at least 15% of the participating students were two or more years below grade level in reading.

Student Writing Learning Outcomes

We analyzed students' historical essays using three writing measures, focusing on their ability to write historically, the overall quality of their writing, and the length of their essays both before and after the yearlong curriculum intervention.

Historical Writing

This dependent variable served as a measure of specific aspects of historical thinking evident in writing and was based on an analytic trait rubric developed by Monte-Sano (2010) that focused on four specific aspects of historical reasoning—substantiation, perspective recognition, contextualization, and rebuttal—and resulted 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. Rebuttal proffered opposing side claims. These can be presented but not addressed, or be addressed with simple to elaborated counterclaims, or critique. We share excerpts from three levels of a rubric for each historical writing trait analyzed alongside excerpts from students' essays to illustrate the scoring process and results (see Tables 3–6 for a description of each trait and excerpts from students' essays).

[COMP: Please insert Tables 3, 4, 5, and 6.]

We taught two or three pairs of raters in each year of the study to use the analytic trait rubric, asking them to consider one trait at a time and talk through distinctions in scores. The raters scored the entire set of 2,058 pre- and posttest essays, working on subsets of data from each year of the project separately, and achieved satisfactory reliability for each analytic trait (Spearman's *r* for perspective = .94 for year 2 and .96 for year 3; Spearman's *r* for substantiation = .89 for year 2 and .92 for year 3; Spearman's *r* for contextualization = .89 for year 2 and .94 for year 3; Spearman's *r* for rebuttal = .92 for year 2 and .94 for year 3). The separate scores were combined, and the summary score was standardized to have a mean of 0 and a standard deviation of 1. Students' standardized pretest abilities to write historical arguments ranged from −2.18 to 2.81 in both the full sample and the treatment sample. Standardized posttest scores ranged from −2.44 to 2.02 in both the full sample and 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. 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 (see Table 7 for descriptors and examples of student writing that exemplifies different levels of quality). 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. Three pairs of raters scored the complete set of essays in each year, with .88 inter-rater agreement in year 2 and .91 in year 3 (Spearman's *r*). The measure was standardized to have a mean of 0 and a standard deviation of 1. Standardized pretest holistic quality scores ranged from −2.48 to 3.40 in both the full sample and the treatment sample. Standardized posttest scores ranged from −2.28 to 2.05 in both the full sample and the treatment sample.

[COMP: Please insert Table 7.]

Essay Length

This dependent variable consisted of the number of words written. Although not a measure of essay quality, we consider length an indicator of automaticity or general ease in writing (Kobrin, Deng, & Shaw, 2007; Quinlan, 2004), which has been shown to be positively correlated with overall writing ability (Gregg, Coleman, Davis, & Chalk, 2007). For the purposes of this study, the length comprised all words that represented a spoken word regardless of spelling. Scoring conventions included counting "nine o'clock p.m." as three words, "Mil I tary" as one word, "United States" as two words, and "1863–1865" as three words. Independent raters scored all essays. Independent readers in year 2 counted a random sample of 100 papers, with adequate reliability (Pearson's $r = .99$ in year 2); in addition, all papers were counted in year 3 with the same degree of reliability (Pearson's $r = .99$). This measure was standardized to have a mean of 0 and a standard deviation of 1. Standardized pretest essay length ranged from −1.66 to 8.71 in both the full sample and the treatment sample. Standardized posttest scores ranged from −1.88 to 6.06 in the full sample and from −1.88 to 4.77 in the treatment sample.

Student Demographic Variables

We conducted preliminary analyses with several variables to control for students' characteristics, which allowed us to explore the influence of their background and incoming reading and writing abilities, using information from the school district and information from a standardized writing test (the Test of Written Language, fourth edition; TOWL–4). A description of these variables is as follows: We included students' gender with a dichotomized variable, such that 0 represented females and 1 represented males. We included indicators of racial status with variables (White, Hispanic, Native American, Asian/Pacific Islander, and Other), each dichotomized such that 1 represented the respective race and 0 represented "other race." We also included dichotomous

indicators of whether the student was an English learner (EL) or had an IEP. To capture prior achievement, we included the standardized version of TOWL–4 scores. We also included two dichotomous variables to indicate whether students had reached proficient or advanced levels on the state's reading assessment in the spring prior to the intervention.

In later descriptive analyses, we explored the extent to which students completed lesson components that corresponded with disciplinary activities (reading, planning, and writing) by developing an elaborate coding procedure to tabulate each student's attempt at completing critical lesson components. Prior to running the analyses, pre- and postintervention historical reasoning, writing quality, and essay length scores were standardized to have a mean of 0 and a standard deviation of 1 across the full sample.

Teachers' Fidelity to the Intervention

We tracked teachers' fidelity to central aspects of cognitive apprenticeship by attending to four factors for every classroom observation: (a) building an understanding of the historical reading and writing strategies through modeling and collaborative practice; (b) promoting independence in students' comprehension and use of the strategies through feedback about student learning and fading scaffolds when appropriate; (c) building students' historical/topical knowledge; and (d) promoting a positive learning environment for student learning to take place, whether through the use of classroom routines or by adapting instruction to meet the needs of specific groups of students or events (e.g., prompting struggling readers, a shorter class period because of testing).

To determine whether teachers implemented the curriculum intervention as planned, we developed observation protocols for each lesson in each of the six investigations. Results from this tool then helped us evaluate the effects of differing levels of fidelity. Scores represented the degree to which teachers adhered to core constructs of the intervention, according to an observer. 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 the elements 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 strategy instruction that were instrumental in helping students gain independence in their learning.

By looking at the degree to which teachers implement the core components of an intervention, we get more data on whether these components are associated with learning outcomes (O'Donnell, 2008). Additionally, 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.

Students' Fidelity to the Intervention

O'Donnell (2008) suggested that researchers should look beyond what teachers do, when an intervention is implemented, to examine the role of students, such as by the degree to which they complete their lessons, when determining the overall effects of a curriculum intervention. In our study, this variable represents the average work completed per class on three separate lesson activities related to (a) annotating sources, referred to hereafter as reading, in our dependent measures; (b) evaluating and selecting content before composing, referred to as planning; and (c) the number of paragraphs that students were able to write during the third day of each lesson, hence, writing.

Analytic Measures and Statistical Methods

Because we nested students within teachers, we used hierarchical linear modeling. We used a series of two-level random intercept models, with students at level 1 and teachers at level 2, to examine the effects of participating in the disciplinary writing curriculum intervention on three aspects of students' disciplinary writing skills: historical reasoning, writing 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).

At level 1, we modeled the disciplinary writing skills of student *i* with teacher *j* as a function of a vector of student characteristics and random student error (e_{ij}):

[COMP: In the equation below, please break turnovers before a + and hang them after the =.]

$$Y_{ij} = \beta_{0j} + \beta_{1j}(\text{male})_{ij} + \beta_{2j}(\text{white})_{ij} + \beta_{3j}(\text{Hispanic})_{ij} + \beta_{4j}(\text{Native American})_{ij} + \beta_{5j}(\text{Asian})_{ij} + \beta_{6j}(\text{other race})_{ij} + \beta_{7j}(\text{IEP})_{ij} + \beta_{8j}(\text{EL})_{ij} + \beta_{9j}(\text{TOWL–4 grammar})_{ij} + \beta_{10j}(\text{TOWL–4 story construction})_{ij} + \beta_{11j}(\text{proficient})_{ij} + \beta_{12j}(\text{advanced})_{ij} + \beta_{13j}(\text{prescore})_{ij} + e_{ij}$$

where Y_{ij} is a measure of the disciplinary writing skills of student *i* with teacher *j*, and β_{0j} is the average disciplinary writing skills of students of teacher *j*. Student-level variables included a series of dichotomous variables, indicating gender, racial status, whether the student had an IEP,

whether the student was an EL, and preintervention proficiency on a state reading assessment (Maryland School Assessment [MSA] proficiency levels). The continuous student-level variables included the preintervention scores on the TOWL–4 (subtest 6 is related to grammar, and subtest 7 measures story development) and on the pretest for each of the respective outcomes (historical reasoning, writing quality, and essay length). Analyses restricted to the treatment group also included a measure of student fidelity to the intervention. Finally, e_{ij} is the random error or unique effect of student *i* of teacher *j* on the measure of disciplinary writing skills.

At level 2, we modeled the average disciplinary writing skills of students of teacher *j* as a function of participation in the disciplinary writing curriculum intervention and random teacher error (u_j):

$$\beta_{0j} = \gamma_{00} + \gamma_{01}(\text{Treat})_1 + u_j$$

where β_{0j} is the average disciplinary writing skills of students of teacher *j*, γ_{00} is the average disciplinary writing skills of all students in the study, Treat is a dichotomous variable indicating whether the teacher is participating in the disciplinary writing curriculum intervention, γ_{01} is the level 2 coefficient that measures the effect of the disciplinary writing curriculum intervention on average student disciplinary writing skills, and u_j is the random error or unique effect of teacher *j* on students' disciplinary writing skills.

Our models allow the intercept for postintervention disciplinary writing skills to randomly vary between teachers (u_j). All student-level variables were grand mean centered in all analyses. The level 1 intercept, therefore, is the average disciplinary writing skills of students net of differences among teachers in their students' characteristics.

Results

Descriptive Results

Table 1 contains summaries of the variables considered in the analysis, and Table 8 provides correlations between continuous variables for the sample. Students in the intervention group had higher average TOWL–4 scores on the subtest that measured their competence in use of standard English grammar, whereas students in the control group had higher pretest historical writing quality and overall writing quality and longer essay lengths. The posttest scores for each of these measures were also significantly different, with differences favoring the intervention group for all three learning outcomes.

[COMP: Please insert Table 8.]

We conducted *t*-tests and analyses of variance to assess whether pre- and postintervention scores varied by the demographic and academic background variables that we planned to include in our analyses. We found that males and students with IEPs had significantly lower scores on all pre- and postintervention scores. ELs had lower scores on every outcome except postintervention essay length. Asian American students scored significantly higher than white, Hispanic American, and African American students in postintervention overall writing quality. Students who scored in the advanced range on the state standardized reading assessment (MSA) before participating in the intervention had significantly higher scores than students who scored in the proficient range on this test, and who in turn had significantly higher scores than students scoring in the basic range on the MSA, across all pre- and posttest measures of writing. These relationships underscore the importance of controlling for such characteristics in our models so postintervention test scores can more defensibly be attributed to the intervention rather than to characteristics of the students.

Interclass Correlation Coefficient

We first fit a fully unconditional model to examine the variability between the classes in each end-of-year writing score. We determined estimates for random effects, intraclass correlations, and the reliability of the level 1 intercept (β_0), based on a fully unconditional model for each outcome, with results indicating that although about 80% of the variance in students' disciplinary writing skills occurred between students, the initial unconditional models indicated that 17–25% of the variance in the outcomes occurred between teachers. These variance components were significantly different from zero ($p < .001$), and reliability estimates for the intercepts were sufficient for multilevel modeling.

We then fit the level 1 (i.e., student level) models for student characteristics. For the full sample, the student-level models explained 19–24% of the within-teacher variability in the outcomes. Because all level 1 variables are grand mean centered, they control for differences across teachers in the gender and racial compositions of their students, as well as the proportion of students with IEPs, the proportion of ELs, average scores on the two standardized measures of writing on the TOWL–4 (i.e., on a test of grammatical competence and ability to write a story), the proportion of students with advanced and proficient scores on the MSA, and average preintervention scores on the outcomes. These level 1 variables explained 40% of the between-

teacher variability in writing quality scores, although they explained less than 10% of the variation between teachers in historical reasoning and essay length. After controlling for student characteristics, variance components for all three outcomes remained significantly different from zero.

Multilevel Models

Finally, we fit a level 2 model for the curriculum effects. Table 9 presents the two-level, fully conditional models as evidence of the effects of the curriculum intervention on student learning outcomes. Results for the intervention were positive and significant for all three student learning writing outcomes. After controlling for the other variables in the model, our overarching finding was that students in the treatment condition outperformed the control group students. It is important to clarify that in our analyses, we estimated the effects of the curriculum intervention on students' posttest scores, controlling for gender (the referent group for this 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). Basic readers in the treatment condition scored about 1.62 points higher in historical writing and 0.37 points higher in overall writing quality and wrote approximately 60 more words in their essays, relative to the control group students (effect size (ES) = 0.45 on their historical writing, 0.32 on the overall quality of their writing, and 0.60 on the length of their essays).

[COMP: Please insert Table 9.]

After controlling for other variables in the model, we also found the following differences. White students' historical writing quality scores were about a quarter of a standard deviation lower than African American students' on the same measure. Males scored somewhat lower than females on their historical writing quality and essay length; students with IEPs scored slightly lower in their overall writing quality as compared with students without IEPs. The disciplinary writing scores of ELs were no different from their non-EL peers' scores, after controlling for other variables in the model.

Not surprisingly, we found that students who scored in the proficient or advanced range on the state's reading assessment had higher average scores across all writing outcomes, thus demonstrating specific benefits for these students versus struggling readers, as well as differences in the degree to which they benefited from the curriculum intervention. Proficient

readers' scores were 1.58 points higher than basic readers' on historical writing and 0.35 points higher on overall quality. Proficient readers' essays included about 19 more words, on average, than the essays written by basic readers. Among advanced readers, scores averaged 2.53 points higher than basic readers' on historical writing and 0.75 points higher on overall quality; in terms of essay length, advanced readers' essays included about 37 more words than basic readers' essays. When translated into effect sizes, scores for proficient students were 0.44 higher than those of peers who scored in the basic range on the historical writing outcome; moreover, the effect sizes for proficient students were 0.30 higher in overall writing quality and 0.19 for longer essay length. Advanced students outperformed basic students by even greater amounts: This was equivalent to an effect size of 0.70 for advanced students' historical writing quality, 0.65 for their overall writing quality, and 0.37 for their essay length.

In addition, students' preintervention scores were significant predictors of their postintervention scores for the overall writing quality and essay length outcomes, although not for their postintervention measure of historical writing. Specifically, students who scored 1 point higher on writing quality on the pretest scored 0.18 points higher on the posttest, and students who wrote one word more on the pretest wrote 0.39 words more on the posttest. These translate into effect sizes of 0.13 for students who scored one standard deviation higher in writing quality on the pretest, and 0.28 for students scoring one standard deviation higher in essay length prior to the intervention, after controlling for other variables in the model.

Using the random effects for the fully conditional level 1 models reported in Table 9 as a baseline, the intervention explained 30% of the variation across teachers in postintervention historical writing scores, 21% of the variation in postintervention overall writing quality scores, and 34% of the variation in postintervention essay length.

Teacher Fidelity Effects

In addition to examining the overall impact of the curriculum intervention across groups, we wished to learn which parts of our lessons were important to the intended writing outcomes for students who participated in the curriculum intervention. Table 10 provides descriptive statistics regarding the teachers' ability to implement our curriculum intervention with fidelity to each key element and indicates the average percentages across each cohort and for both years, when data are combined. We found these overall results to be encouraging, an indication of the success of our PD efforts across the year.

[COMP: Please insert Table 10.]

Student Fidelity Effects

We cataloged the work completion for all 1,029 participating students, examining each of the 18 days of lessons in the six investigations for the degree to which each student completed reading, planning, and writing activities. This information was then summarized across all 18 days of lessons and all teachers, resulting in overall percentages (reported in Table 11 as percentages for reading and planning and, in the average number of paragraphs written across investigations, for writing). Perhaps most relevant to the students' improvements in writing, data in year 3 show that students planned less and wrote more, overall (in year 2, some students did not finish composing because of the time they spent planning).

[COMP: Please insert Table 11.]

Finally, we share Figure 1, which represents the difference in standardized outcomes for teachers whose students who have high aggregate writing fidelity (i.e., at or above the mean, or about 2.5 paragraphs composed per investigation across both years). These results, adjusted by controlling for students' own writing propensity, show the average contextual (e.g., classroom) effects of writing fidelity above the mean work completion and accounting for beginning ability that influences each writing outcome. In other words, among the group of students who participated in the curriculum intervention, we found that students who were assigned to teachers whose classes wrote more during each investigation had greater gains on the historical writing outcomes than did students with teachers whose classes wrote less on average, even after accounting for students' own writing fidelity.

[COMP: Please insert Figure 1.]

Discussion

The results from this study add to a growing literature on the positive impact of cognitive apprenticeships on middle and high school students' discipline-specific reading and writing (e.g., De La Paz et al., 2014; Nokes, Dole, & Hacker, 2007; Reisman, 2012). Although tentative by virtue of design, the data presented here come from and are representative of a large number of students and suggest that when delivered with fidelity, our curriculum intervention and PD resulted in improved historical writing and general argument writing for diverse learners, especially in comparison with writing from students who were in eighth-grade classes where the program was not provided. Our results also indicate the success of our cognitive apprenticeship

approach for readers at higher proficiency levels and those who struggled academically. Finally, we learned how teachers' actions influenced the impact of the curriculum intervention.

Writing Outcomes

In contrast to our prior results, in the current investigation, we saw significant growth not only in students' historical writing but also in their general argument writing. 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 in the historical writing of students who received the curriculum intervention (1.62 points, equivalent to an effect size of 0.45 for struggling readers whose teachers implemented the curriculum intervention), their overall writing quality (0.37 points, ES = 0.32), and their ease in writing more text (60 words, ES = 0.60), after accounting for variation in students' prior achievement, gender, ethnicity, and differences in these characteristics across classrooms. These results are significant in part for overcoming a limitation noted in an earlier version of our curriculum intervention regarding gains in students' overall writing quality (i.e., the persuasiveness of their argument and its overall organization; De La Paz et al., 2014).

Thus, in this study, students who engaged in our curriculum intervention improved in their ability to write disciplinary arguments, in the quality of their writing, and in their general fluency in producing written text, demonstrating that changes we made to the curriculum intervention actually improved other aspects of students' writing proficiency. This finding suggests that to achieve proficiency in aspects of disciplinary literacy, educators cannot—and need not—leave general literacy aside. This suggests that students can learn discipline-specific and general forms of writing in tandem, regardless of their incoming skills. In other words, students may not need to learn general argument writing before they learn to write historical arguments.

Fidelity of Implementation

Because fidelity of implementation was generally high across teachers, there was not enough variation to study the relative importance of core curriculum components. This was despite our finding that teachers began the year with different levels of disciplinary understandings and historical thinking practices, which we determined by asking them to complete a questionnaire that was designed to measure their pedagogical content knowledge. Briefly, although details are reported elsewhere (Monte-Sano et al., 2014b), we asked teachers to analyze students' work and

devise instructional responses to support students' continued learning of historical reading and writing. We believe that teachers' success in implementing the curriculum and core constructs of cognitive apprenticeship indicates that it is possible to provide instruction in both historical writing and general argument writing at the same time. Not only that, but it is possible to learn to teach disciplinary and general literacy simultaneously. Participating teachers were not accustomed to teaching analytical thinking with primary sources, nor were they familiar with teaching reading and writing. Yet, our fidelity of implementation analyses indicated that during this program, teachers were able to do both given the support of the curriculum intervention and PD.

High teacher fidelity not only confirmed treatment validity but also allowed us to observe more directly the relationship between student fidelity and writing outcomes. When a teacher's class of students wrote an average of about 2.5 paragraphs of the five paragraphs that were expected in each investigation, learning outcomes were higher compared with the outcomes of students in classes that wrote less than 2.5 paragraphs per investigation, on average. In classes where students wrote an average of 2.5 or more paragraphs per investigation, students scored 1.00 points higher in historical writing and 0.48 points higher in overall holistic quality and wrote 43 words more in their essays, relative to students in intervention group classes where students wrote less than 2.5 paragraphs on average. This translates into effect sizes of 0.28 for historical writing, 0.41 for overall holistic quality, and 0.43 for the length of their papers. These effects are realized after taking into consideration the effects of the intervention and characteristics associated with the students' own abilities and work completion. Student fidelity results indicated that writing a whole essay helps students develop historical and general argument writing more so than writing smaller pieces. In other words, the more practice students had in writing complete essays throughout the intervention, the better able they were to master both kinds of writing.

Limitations

The findings reported in this investigation are not without limitations. We first acknowledge that we were not permitted to randomly assign teachers to conditions. The broad purpose of our funded work was to develop and refine an intervention for struggling adolescent readers. Moreover, although we learned about teachers who implemented our curriculum intervention, we know little about the teachers who composed our control groups or the instruction they delivered

to students in their classrooms. We know that some teachers who were unable to participate initially joined us in the current study, however, and that some teachers could not join the intervention group because of other school-related obligations. Thus, although we are unable to determine the extent to which teachers in the control group were similar to those who joined the experimental condition, we also do not suspect that there were major differences in terms of their teaching preparation or years of experience. Moreover, some would contend that differences in teacher background do not necessarily translate into differences in student learning outcomes (Yoon, Duncan, Lee, Scarloss, & Shapley, 2007).

Finally, although we did not observe control teachers more than occasionally, the cooperating district had initiated a PD program focused on disciplinary literacy in history that involved sharing primary document–based lessons two times per year with social studies teachers. In addition, all district teachers were held accountable for using the same pacing guide for instruction that listed the information to be covered by various points in the school year.

Conclusions

Our findings have implications for practitioners, as 43 states have now adopted the Common Core State Standards and national organizations in both the United States and Canada have called for increased attention to viewing history as a discipline with standards related to the development of historical reasoning (e.g., the C3 Framework for Social Studies State Standards; National Council for the Social Studies, 2013). Writing demands that students read and analyze texts, organize their thoughts, and compose essays, keeping their reading and analysis in mind. The results from this study show that cultivating students' historical writing and general argument writing practices in the classroom is possible when using a cognitive apprenticeship model to teach integrated reading and writing strategies that makes explicit links among general literacy, disciplinary literacy, and content learning, along with supportive and sustained teacher PD. Although there are multiple facets to historical writing, including concepts such as historical significance, continuity and change, cause and consequence, historical perspectives, and moral dimensions of history (cf. Seixas, 2006), the results of this study demonstrate that writing advanced historical arguments is not restricted to college or advanced high school students and is within reach for young adolescents, including those who struggle with reading. Given the structure of this study, we wonder how much PD is necessary for treatment effects with students.

Our findings have implications for researchers as well, especially for those interested in developing and modifying cognitive apprenticeship models of instruction for disciplinary reading and writing tasks. Future research might explore how teachers, learning environments, materials, and tools might be developed to support students as they try to shift their conceptual knowledge of history or grapple with more challenging historical tasks and writing genres that more closely approximate the work of older students (e.g., college age) and historians. We believe that older academically diverse learners are likely to benefit from similar, systematic approaches to instruction that emphasize the flexible coordination of historical reading, thinking, and writing with content learning, as learners attempt to regulate underlying cognitive processes that are specific to historical writing. It remains to be seen whether there are limits to this approach to instruction as disciplinary literacy demands become more complex or whether cognitive apprenticeships remain a viable means for helping even advanced learners gain mastery of more complex historical writing.

Notes

The Institute of Education Sciences of the U.S. Department of Education supported the research reported here through a grant (R305A090153) awarded to the University of Maryland, College Park. The opinions expressed herein are those of the authors and do not represent views of the Institute of Education Sciences or the U.S. Department of Education. We are grateful to the teachers and students who made this research possible.

References

Bain, R. (2006). Rounding up unusual suspects: Facing the authority hidden in the history classroom. *Teachers College Record*, *108*(10), 2080–2114.

Bain, R. (2012). Using disciplinary literacy to develop coherence in history teacher education: The Clinical Rounds Project. *The History Teacher*, *45*(4), 513–532.

Bangert-Drowns, R.L., Hurley, M.M., & Wilkinson, B. (2004). The effects of school-based writing-to-learn interventions on academic achievement: A meta-analysis. *Review of Educational Research*, *74*(1), 29–58. doi:10.3102/00346543074001029

Beaufort, A. (2004). Developmental gains of a history major: A case for building a theory of disciplinary writing expertise. *Research in the Teaching of English*, *39*(2), 136–185.

Brown, J.S., Collins, A., & Duguid, P. (1989). Situated cognition and the culture of learning. *Educational Researcher*, *18*(1), 32–42. doi:10.3102/0013189X018001032

Carter, M., Ferzli, M., & Wiebe, E.N. (2007). Writing to learn by learning to write in the disciplines. *Journal of Business and Technical Communication*, *21*(3), 278–302. doi:10.1177/1050651907300466

Century, J., Rudnick, M., & Freeman, C. (2010). A framework for measuring fidelity of implementation: A foundation for shared language and accumulation of knowledge. *American Journal of Evaluation*, *31*(2), 199–218. doi:10.1177/1098214010366173

Coffin, C. (2006). *Historical discourse: The language of time, cause, and evaluation*. London, UK: Continuum.

Collins, A., Brown, J.S., & Holum, A. (1991). Cognitive apprenticeship: Making thinking visible. *American Educator*, *15*(3), 6–11, 38–46.

Conley, M.W. (2012). Foregrounding the disciplines for teacher preparation in secondary literacy. *Journal of Adolescent & Adult Literacy*, *56*(2), 141–150. doi:10.1002/JAAL.00115

Cuban, L. (1991). History of teaching in social studies. In J.P. Shaver (Ed.), *Handbook of research on social studies teaching and learning* (pp. 197–209). New York, NY: Macmillan.

De La Paz, S. (2005). Effects of historical reasoning instruction and writing strategy mastery in culturally and academically diverse middle school classrooms. *Journal of Educational Psychology*, *97*(2), 139–156. doi:10.1037/0022-0663.97.2.139

De La Paz, S., & Felton, M.K. (2010). Reading and writing from multiple source documents in history: Effects of strategy instruction with low to average high school writers. *Contemporary Educational Psychology*, *35*(3), 174–192. doi:10.1016/j.cedpsych.2010.03.001

De La Paz, S., Felton, M., Monte-Sano, C., Croninger, R., Jackson, C., Deogracias, J.S., & Hoffman, B.P. (2014). Developing historical reading and writing with struggling adolescent readers: Implementation and learning outcomes. *Theory & Research in Social Education*, *42*(2), 228–274. doi:10.1080/00933104.2014.908754

Draper, R.J. (2008). Redefining content-area literacy teacher education: Finding my voice through collaboration. *Harvard Educational Review*, *78*(1), 60–83. doi:10.17763/haer.78.1.k104608143l205r2

Englert, C.S., Raphael, T.E., Anderson, L., Stevens, D.D., & Fear, K.L. (1991). Making writing strategies and self-talk visible: Cognitive strategy instruction in writing in regular and special education classrooms. *American Educational Research Journal*, *28*(2), 337–372. doi:10.3102/00028312028002337

Greenleaf, C.L., Schoenbach, R., Cziko, C., & Mueller, F. (2001). Apprenticing adolescent readers to academic literacy. *Harvard Educational Review*, *71*(1), 79–130. doi:10.17763/haer.71.1.q811712577334038

Gregg, N., Coleman, C., Davis, M., & Chalk, J.C. (2007). Timed essay writing: Implications for high-stakes tests. *Journal of Learning Disabilities*, *40*(4), 306–318. doi:10.1177/00222194070400040201

Grossman, P., Hammerness, K., & McDonald, M. (2009). Redefining teaching, re-imagining teacher education. *Teachers and Teaching: Theory and Practice*, *15*(2), 273–289. doi:10.1080/13540600902875340

Harris, K., & Graham, S. (1996). *Making the writing process work: Strategies for composition and self-regulation* (2nd ed.). Cambridge, MA: Brookline.

Heller, R. (2010). In praise of amateurism: A friendly critique of Moje's "Call for Change" in secondary literacy. *Journal of Adolescent & Adult Literacy*, *54*(4), 267–273. doi:10.1598/JAAL.54.4.4

Hexter, J.H. (1971). *The history primer*. New York, NY: Basic.

Jitendra, A.K., & Gajria, M. (2011). Main idea and summarization instruction to improve reading comprehension. In R.E. O'Connor & P.F. Vadasy (Eds.), *Handbook of reading interventions* (pp. 198–218). New York, NY: Guilford.

Kiuhara, S., Graham, S., & Hawken, L. (2009). Teaching writing to high school students: A national survey. *Journal of Educational Psychology*, *101*(1), 136–160. doi:10.1037/a0013097

Kobrin, J., Deng, H., & Shaw, E.J. (2007). Does quantity equal quality? The relationship between length of response and scores on the SAT essay. *Journal of Applied Testing Technology*, *8*(1), 1–15.

Le Bigot, L., & Rouet, J.F. (2007). The impact of presentation format, task assignment, and prior knowledge on students' comprehension of multiple online documents. *Journal of Literacy Research*, *39*(4), 445–470. doi:10.1080/10862960701675317

McCoach, D.B. (2010). Hierarchical linear modeling. In G.R. Hancock & R.O. Mueller (Eds.), *Quantitative methods in the social and behavioral sciences: A guide for researchers and reviewers* (pp. 123–140). New York, NY: Routledge.

Meyer, B.J.F., Brandt, D.M., & Bluth, G.J. (1980). Use of top-level structure in text: Key for reading comprehension of ninth-grade students. *Reading Research Quarterly*, *16*(1), 72–103. doi:10.2307/747349

Meyer, B.J.F., & Freedle, R.O. (1984). Effects of discourse type on recall. *American Educational Research Journal*, *21*(1), 121–143. doi:10.3102/00028312021001121

Moje, E.B. (2008). Foregrounding the disciplines in secondary literacy teaching and learning: A call for change. *Journal of Adolescent & Adult Literacy*, *52*(2), 96–107. doi:10.1598/JAAL.52.2.1

Monte-Sano, C. (2008). Qualities of historical writing instruction: A comparative case study of two teachers' practices. *American Educational Research Journal*, *45*(4), 1045–1079. doi:10.3102/0002831208319733

Monte-Sano, C. (2010). Disciplinary literacy in history: An exploration of the historical nature of adolescents' writing. *Journal of the Learning Sciences*, *19*(4), 539–568. doi:10.1080/10508406.2010.481014

Monte-Sano, C. (2011). Beyond reading comprehension and summary: Learning to read and write by focusing on evidence, perspective, and interpretation. *Curriculum Inquiry*, *41*(2), 212–249. doi:10.1111/j.1467-873X.2011.00547.x

Monte-Sano, C., De La Paz, S., & Felton, M. (2014a). Implementing a disciplinary literacy curriculum for US history: Learning from expert middle school teachers in diverse classrooms. *Journal of Curriculum Studies*, *46*(4), 540–575. doi:10.1080/00220272.2014.904444

Monte-Sano, C., De La Paz, S., & Felton, M. (2014b). *Reading, thinking, and writing about history: Teaching argument writing to diverse learners in the Common Core classroom, grades 6–12*. New York, NY: Teachers College Press.

National Center for Education Statistics. (2015). *The Nation's Report Card: 2014 U.S. history, geography, and civics at grade 8* (NCES 2015-112). Washington, DC: National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education.

National Council for the Social Studies. (2013). *The College, Career, and Civic Life (C3) Framework for Social Studies State Standards: Guidance for enhancing the rigor of K–12 civics, economics, geography, and history*. Silver Spring, MD: Author.

National Governors Association Center for Best Practices & Council of Chief State School Officers. (2010). *Common Core State Standards for English language arts and literacy in history/social studies, science, and technical subjects*. Washington, DC: Authors.

Nokes, J.D. (2010). Observing literacy practices in history classrooms. *Theory & Research in Social Education*, *38*(4), 515–544. doi:10.1080/00933104.2010.10473438

Nokes, J.D. (2013). *Building students' historical literacies: Learning to read and reason with historical texts and evidence*. New York, NY: Routledge.

Nokes, J.D., Dole, J.A., & Hacker, D.J. (2007). Teaching high school students to use heuristics while writing historical texts. *Journal of Educational Psychology*, *99*(3), 492–504. doi:10.1037/0022-0663.99.3.492

O'Donnell, C.L. (2008). Defining, conceptualizing, and measuring fidelity of implementation and its relationship to outcomes in K–12 curriculum intervention research. *Review of Educational Research*, *78*(1), 33–84. doi:10.3102/0034654307313793

Quinlan, T. (2004). Speech recognition technology and students with writing difficulties: Improving fluency. *Journal of Educational Psychology*, *96*(2), 337–346. doi:10.1037/0022-0663.96.2.337

Ragland, R.G. (2007). Changing secondary teachers' views of teaching American history. *The History Teacher*, *40*(2), 219–246.

Reisman, A. (2012). Reading like a historian: A document-based history curriculum intervention in urban high schools. *Cognition and Instruction*, *30*(1), 86–112. doi:10.1080/07370008.2011.634081

Rouet, J.F., Britt, M.A., Mason, R.A., & Perfetti, C.A. (1996). Using multiple sources of evidence to reason about history. *Journal of Educational Psychology*, *88*(3), 478–493. doi:10.1037/0022-0663.88.3.478

Seixas, P. (2006). *Benchmarks of historical thinking: A framework for assessment in Canada*. Vancouver, CA: Center for the Study of Historical Consciousness. Retrieved from archive.historybenchmarks.ca/documents/benchmarks-historical-thinking-framework-assessment-canada

Shanahan, T., & Shanahan, C. (2012). What is disciplinary literacy and why does it matter? *Topics in Language Disorders*, *32*(1), 7–18. doi:10.1097/TLD.0b013e318244557a

Smith, J., & Niemi, R. (2001). Learning history in school: The impact of course work and instructional practices on achievement. *Theory & Research in Social Education*, *29*(1), 18–42. doi:10.1080/00933104.2001.10505928

Stahl, S., Hynd, C., Britton, B., McNish, M., & Bosquet, D. (1996). What happens when students read multiple source documents in history? *Reading Research Quarterly*, *31*(4), 430–456. doi:10.1598/RRQ.31.4.5

Stevens, R., Wineburg, S., Herrenkohl, L.R., & Bell, P. (2005). Comparative understanding of school subjects: Past, present, and future. *Review of Educational Research*, *75*(2), 125–157. doi:10.3102/00346543075002125

Torgesen, J.K., Houston, D.D., Rissman, L.M., Decker, S.M., Roberts, G., Vaughn, S., ... Lesaux, N. (2007). *Academic literacy instruction for adolescents: A guidance document from the Center on Instruction*. Portsmouth, NH: Center on Instruction, RMC Research Corporation.

VanSledright, B. (2002). *In search of America's past: Learning to read history in elementary school*. New York, NY: Teachers College Press.

Westhoff, L.M. (2009). The use of primary sources in teaching history. In R.G. Ragland & K.A. Woestman (Eds.), *The Teaching American History project: Lessons for history educators and historians* (pp. 62–77). New York, NY: Routledge.

Wiley, J., & Voss, J.F. (1999). Constructing arguments from multiple sources: Tasks that promote understanding and not just memory for text. *Journal of Educational Psychology*, *91*(2), 301–311. doi:10.1037/0022-0663.91.2.301

Wilson, S. (Ed.). (2009). *Teacher quality* [Education policy white paper]. Washington, DC: National Academy of Education.

Wineburg, S. (1991). Historical problem solving: A study of the cognitive processes used in the evaluation of documentary and pictorial evidence. *Journal of Educational Psychology*, *83*(1), 73–87. doi:10.1037/0022-0663.83.1.73

Wineburg, S., & Martin, D. (2009). Tampering with history: Adapting primary sources for struggling readers. *Social Education*, *73*(5), 212–216.

Yoon, K.S., Duncan, T., Lee, S.W.-Y., Scarloss, B., & Shapley, K. (2007). *Reviewing the evidence on how teacher professional development affects student achievement* (REL 2007–No. 033). Washington, DC: Regional Educational Laboratory Southwest, National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.

Young, K.M., & Leinhardt, G. (1998). Writing from primary documents: A way of knowing in history. *Written Communication*, *15*(1), 25–68. doi:10.1177/0741088398015001002

Submitted September 29, 2015

Final revision received February 25, 2016

Accepted March 2, 2016

[COMP: Please set the authors' names per RRQ's design.]

SUSAN DE LA PAZ (corresponding author) is an associate professor in the College of Education at the University of Maryland, College Park, USA; e-mail sdelapaz@umd.edu. Her research examines writing and argumentation, and the application of a cognitive apprenticeship model to meet disciplinary literacy goals with academically diverse adolescents.

CHAUNCEY MONTE-SANO is an associate professor in the School of Education at the University of Michigan, Ann Arbor, USA; e-mail cmontesa@umich.edu. Her research examines how history students learn to reason with evidence in writing and how their teachers learn to teach such historical thinking.

MARK FELTON is a professor in the College of Education at San Jose State University, California, USA; e-mail mark.felton@sjsu.edu. His research focuses on argumentative discourse and its role in shaping students' reasoning, evidence use, and literacy in school settings.

ROBERT CRONINGER is an associate professor in the College of Education at the University of Maryland, College Park, USA; e-mail croninge@umd.edu. His research examines how

education policies and practices affect the educational experiences of students from historically disadvantaged and underserved backgrounds, including how the social organization of schools and different instructional practices affect the distribution of educational opportunities.

CARA JACKSON is the assistant director of research and evaluation at Urban Teachers, Baltimore, Maryland, USA; e-mail cara.jackson@urbanteachers.org. Her research focuses on efforts to improve the quality of teacher preparation and evaluation.

KELLY WORLAND PIANTEDOSI is a teacher in the District of Columbia Public Schools, Washington, DC, USA; kelly.worland@dc.gov. Her research interests include writing instruction and assessment in the elementary grades.

Supporting Information

Additional supporting information may be found in the online version of this article:

- Appendix A: IREAD Poster: A Scaffold for Historical Reading
- Appendix B: How to Write Your Essay: A Scaffold for Historical Writing

FIGURE 1

Effects of Student Writing Fidelity

TABLE 1

Participant Characteristics

Variable	Full sample (*N* = 1,029)		Control group (*N* = 384)		Intervention group (*N* = 645)		Significance
	Mean	*SD*	Mean	*SD*	Mean	*SD*	
Male	49.5%		51.6%		48.2%		
White	5.2%		6.0%		4.8%		
Hispanic	21.8%		22.7%		21.2%		
African American	72.2%		70.6%		73.2%		
Native American	10.3%		8.9%		11.2%		
Asian/Pacific Islander	5.2%		7.0%		4.2%		*
Other race	7.0%		7.6%		6.7%		
Individualized Education Plan	6.9%		6.3%		7.3%		
English learner	5.3%		7.3%		4.0%		*

TOWL–4 grammar	10.520	3.004	10.194	2.786	10.713	3.111	**
TOWL–4 story development	11.077	3.359	11.263	3.029	10.967	3.537	
Basic MSA	26.1%	0.439	24.3%	0.429	27.1%	0.445	
Proficient MSA	46.8%	0.499	47.5%	0.500	46.4%	0.499	
Advanced MSA	25.8%	0.438	27.7%	0.448	24.7%	0.431	
Pre historical reasoning	6.542	3.007	7.141	2.905	6.185	3.012	***
Pre writing quality	3.110	0.850	3.197	0.766	3.058	0.893	*
Pre essay length	120.420	72.369	129.381	68.896	115.103	73.893	**
Post historical reasoning	8.767	3.590	7.792	2.903	9.344	3.827	***
Post writing quality	3.638	1.154	3.413	0.835	3.772	1.291	***
Post essay length	188.672	100.380	154.152	81.946	209.126	104.666	***

Note. MSA = Maryland School Assessment; *SD* = standard deviation; TOWL–4 = Test of Written Language, fourth edition.

*$p < .05$. **$p < .01$. ***$p < .001$.

[COMP: Please hang turnovers in bulleted lists.]

TABLE 2

Sequence of the Historical Investigations and the Introduction of Disciplinary Practices

Historical investigation	Disciplinary practices introduced
#1, Lexington Green (days 1–3): "Who fired the first shot at Lexington Green?"	• Historical reading: Sourcing primary sources • Historical writing: Composing a claim
#2, Shays' Rebellion (days 4–6): "Were Daniel Shays and his followers rebels or freedom fighters?"	• Historical reading: Contextualizing primary sources • Historical writing: Identifying the components and structure of a historical argument
#3, Alien and Sedition Acts (days 7–9): "Did the Alien and Sedition Acts violate the U.S. Constitution?"	• Historical reading: Considering authors' evidence • Historical writing: Planning an essay
#4, Indian Removal (days 10–12): "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?"	• Historical reading: Discussing and evaluating evidence • Historical writing: Composing a full essay
#5, Abolitionism (days 13–15): "What was the most promising path toward freeing slaves in the U.S. before the Civil War: nonviolence ("moral persuasion") or more aggressive action?"	• Students set goals to read, analyze, plan, and compose with greater independence.

#6, Mexican–American War (days 16–18): "Was the U.S. justified in going to war with Mexico in 1846?"	• Students integrate reading analysis, planning, and composing independently.

TABLE 3

Perspective Recognition Criterion for the Historical Writing Rubric and Corresponding Excerpts of Student Essays

Score	Rubric descriptor	Student essay excerpt
0	The student presents evidence from documents as student's own perspective (e.g., reports as though factual, does not mention documents or where information came from).	"Most of the slaves went to school if they were slaves or freed. Some white people wanted to close schools for Africans, but the Africans refused too."
2	The student mentions the author (e.g., "According to Lynch..."; "The author says...").	"The author states that 'many grown people want to know how to read.' This shows that...."
4	In using evidence/explanation to support an argument, the writer (a) evaluates the author's perspectives (e.g., discusses reliability, trustworthiness, or credibility) OR (b) evaluates the author's position as a reporter.	The student quotes a source, explains it, and then writes, "This is also reliable because it is the voice of the African Americans and had 24 signatures." OR "Captain Hamilton is very reliable because he actually was their to witness some things and was a soldier in the U.S. army."

TABLE 4

Contextualization Criterion for the Historical Writing Rubric and Corresponding Excerpts of Student Essays

Score	Rubric descriptor	Student essay excerpt
0	(a) No context is mentioned, or inaccurate contextual information overwhelms accurate contextual information. OR (b) The student uses anachronisms (e.g., makes a chronological mistake, uses information from another time period without noting the different era) or generalizations not specific to the time period.	"The African Americans were free because of Reconstruction. Reconstruction started the Civil War…."
2	The student includes factual details about the context of the documents themselves (e.g., mentions the time, place, or audience of the	"One reason is a quote from an excerpt adapted from a letter written by Captain C.M. Hamilton in 1866 to the Office of the Adjutant General in

	documents). This information might come from the documents, headnotes, background information, or source lines.	Washington, D.C....."
4	(a) The writer notes relationships between historical events or situates the documents or argument in the historical setting. OR (b) The writer demonstrates an understanding of the time period (e.g., the norms and beliefs of the Reconstruction era) and goes beyond the specific information in the documents.	"Supposeably James Lynch is seeing that the African Americans are showing that they are free, and little by little are losing 'fear' they once had. This is non reliable because, it was written in the year 1865. That was when they were just starting off. Of course it was going to be 'easy.' But in the year 1867 (2 years after) is when others envied African Americans and got meaner. This was written 2 years 'before,' it's old news. If James Lynch were to go see them now, who knows what he would say."

TABLE 5

Substantiation Criterion for the Historical Writing Rubric and Corresponding Excerpts of Student Essays

Score	Rubric descriptor	Student essay excerpt
0	(a) No position or claim OR (b) no support	"I agree with Side A because it is right. Side B says that."
2	(a) The position is clear. There is clear and relevant support in the essay, but evidence is not drawn from the documents. OR (b) The position is clear. There is clear and relevant support drawn from the documents without explanation.	"Yes, because in Document 1 it say that the African kids went to school. And Charlotte said that she never saw children so eager to learn the alphabet, [the] majority of students learned quickly, and the older one worked in the fields [from] early mornings to 11:00 or 12:00."
4	The position is clear. Evidence is clearly drawn from the documents to support a claim, the link to the claim is clearly established, AND the strength of the evidence or reasoning is evaluated to add support to the claim. (*Note:* In evaluating, the student must not only make a judgment but also share his or her reason for that judgment (i.e., the evaluation must be explained, or the student must show his or her reasoning).)	"After reading information from both sides, I feel African-Americans were not free after the Civil War. Document B says, 'But when we are 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.' This quote is saying African Americans are still being 'harassed' and nobody is doing anything to help them. This point makes sense because the whites didn't necessarily feel comfortable with

		'negroes' around them."

TABLE 6

Rebuttal Criterion for the Historical Writing Rubric and Corresponding Excerpts of Student Essays

Score	Rubric descriptor	Student essay excerpt
0	No mention, acknowledgment, or recognition of opposing sides	In one two-page essay, a student described the perspective of Mr. Lynch, an African American minister, regarding the question. The student never acknowledged an opposing perspective.
2	(a) Opposing sides are presented and clearly distinguished or juxtaposed but are not drawn from the documents. (They may or may not be elaborated on). OR (b) Opposing sides are drawn from the documents and are distinguished or acknowledged but not elaborated on.	"Yes they was free but...the colords would get beet shot knock out could for now reason and the police didn't do nothing about it."
4	Opposing sides are presented and drawn from the documents. In addition, opposing sides are elaborated on. There is an explicit rebuttal, critique of evidence, or reconciliation of opposing views. The student may not take one side in the end but demonstrates the ability to critique at least one side.	"In some ways they were free. Like they could go to school and to church. But the truth was that they were not free. They could go to school but people were predigest against them. Like in the letter Hamilton wrote in 1866. This letter tells us how they were attacking a schoolhouse....So in a way they were free but still they weren't free from the attacks and the hatered...." OR "The colored people were free but were not treated like they were. These are 24 people that have these promblems maybe more. The document A says they have been free which they are but doesn't know how people treating them. I can conclude that the colored people were free but were not getting all the rights they should have gotten."

TABLE 7

Holistic Writing Quality Rubric and Corresponding Excerpts of Student Essays

Score	Rubric descriptor	Student essay excerpt
2	The essay is difficult to follow. The essay addresses the prompt, but the author's	"No they were free because they stopped attacking Americans. After they won the war, they stop

	position may be difficult to determine. There is a confused or incoherent discussion of the subject. The essay lacks the important elements of structure, presents paragraphs and/or sentences nonsequentially or randomly, and lacks transitions and/or topic sentences.	threatening them. Because is there in reason why they should mess with them."
4	The essay is clear but with little development in persuasiveness or structure. The essay clearly addresses the prompt and takes a clear position on the issue, although it may not be explicitly stated. The ideas are coherent and consistent with the points within each paragraph. There is some development of ideas to make them more persuasive. There is little organization beyond the paragraph level. The paragraphs generally progress logically, although they may seem randomly organized. Transitions may be implicit, if present at all.	"African Americans were free after Civil War but they weren't acting as if they were and weren't treated as if they were free. From document 2 this is an example of how they were treated. "Colored men have been knocked down beaten for no reason and yet the police do not notice it at all." As you can tell in that document African Americans are free but to some people they are still seen as slaves and didn't want them to be treated equally. And in document 1 it says: "The colored did not seem to realize that they were free; this was not announced to them." So even in that document it clearly says they were free but wasn't being treated like they were free."
6	The essay is clear, purposeful, persuasive, and well structured. The essay clearly addresses the prompt and takes a clear and explicit position on the issue. The ideas are coherent and consistent and build a persuasive argument. Within paragraphs, there is a clear and purposeful development of ideas, which may even anticipate and address a critical audience. Overall, the essay has a clear and logical structure. The paragraphs are unified and coherent, both internally and /or from paragraph to paragraph. There is evidence of clear transitions and topic sentences.	"After the Civil War, it was questioned whether or not African Americans were truly free. Some believed that they were free and had rights. Others, such as myself, believe that they were not yet genuinely free. According to the letter written by 24 African Americans to the commander of a military district, houses were broken into, shots were fired, and men beaten. Police were not doing anything and the African Americans didn't feel safe. Although this document contains pathos/is emotional at a few points, it is a trustworthy source written by African Americans in need of protection. Also, in the letter from 24 African Americans they explain that all they want is to live in peace and quiet

		obeying the laws. Clearly this is true because they do not try to fight back at the whites that were mistreating them. Instead of violence, they turn to a military commander for help. All they wanted was to have peace. In Lynch's letter to the Relief Association it is said that progress was being made and African Americans were gaining confidence with new rights. Although it may have been accurate at the time, this letter was written over 2 years before the letter by 24 African Americans. During all that time, their rights could have been revoked. The whites may have become tired of their freedom and taken control. So it is not a trustworthy source. If they were free, their rights would have remained the same. So, in conclusion, it is clear that after the Civil War, African Americans were still not completely free. They still were being treated unfairly and unequally by whites. I think that if they were truly free, the rights gained from the war would have remained effective 2 years later."

TABLE 8

Correlations Between Continuous Variables for the Full Sample

Variable	1	2	3	4	5	6	7	8
1. TOWL–4 grammar	—							
2. TOWL–4 story development	.627	—						
3. Pre historical reasoning	.317	.371	—					
4. Post historical reasoning	.306	.310	.201	—				
5. Pre writing quality	.335	.345	.534	.296	—			
6. Post writing quality	.391	.348	.319	.597	.325	—		
7. Pre essay length	.310	.323	.530	.277	.495	.290	—	
8. Post essay length	.289	.281	.234	.614	.261	.627	.354	—

Note. TOWL–4 = Test of Written Language, fourth edition. All correlations are statistically significant at $p < .001$.

[COMP: Plesae shade rows as shown.]

TABLE 9

Effects of the Curriculum Intervention on Student Learning Outcomes for the Full Sample (students *n* = 1,029; teachers *n* = 36)

Variable	Outcome (all standardized)		
	Historical reasoning	Writing quality	Essay length
Intercept	−0.305	−0.213	−0.403
Treatment	0.452***	0.316**	0.602***
Male	−0.153**	−0.032	−0.117*
White	−0.297*	−0.022	−0.062
Hispanic	0.138	−0.013	0.105
Native American	−0.080	−0.058	−0.047
Asian American/Pacific Islander	0.216†	0.151	0.147
Other race	−0.014	0.014	0.021
Individual Education Plan	−0.147	−0.312**	−0.035
English learner	−0.130	−0.100	0.038
Proficient MSA	0.441***	0.302***	0.194**
Advanced MSA	0.703***	0.650***	0.372***
Prescore	0.040	0.133***	0.283***
Variance component			
Intercept	0.110***	0.086***	0.165***
Reliability			
Intercept β_0	0.807	0.772	0.878

Note. MSA = Maryland School Assessment.

†$p < .10$. *$p < .05$. **$p < .01$. ***$p < .001$.

TABLE 10

Percentage Accuracy in Implementing Key Components of the Curriculum Averaged Across Teachers

Variable	Overall (*N* = 22)	Year 2 (*N* = 14)	Year 3 (*N* = 8)
Building understanding	71.30	70.33	73.01
Promoting independence	68.98	74.08	60.05
Building topical knowledge	83.82	82.21	86.64
Learning environment	80.99	82.54	78.29

TABLE 11

Percentage Completion of Key Elements of the Intervention Averaged Across Students

Variable	Overall (*N* = 1,029)	Year 2 (*N* = 384)	Year 3 (*N* = 645)
Reading	77.23	77.69	76.41
Planning	61.94	72.25	43.89
Writing (in paragraphs)	2.42	2.32	2.61