

Written Communication

<http://wcx.sagepub.com/>

Adolescents' Disciplinary Use of Evidence, Argumentative Strategies, and Organizational Structure in Writing About Historical Controversies

Susan De La Paz, Ralph Ferretti, Daniel Wissinger, Laura Yee and Charles MacArthur

Written Communication 2012 29: 412 originally published online 3 October 2012
DOI: 10.1177/0741088312461591

The online version of this article can be found at:
<http://wcx.sagepub.com/content/29/4/412>

Published by:



<http://www.sagepublications.com>

On behalf of:

[Annenberg School for Communication and Journalism](#)

Additional services and information for *Written Communication* can be found at:

Email Alerts: <http://wcx.sagepub.com/cgi/alerts>

Subscriptions: <http://wcx.sagepub.com/subscriptions>

Reprints: <http://www.sagepub.com/journalsReprints.nav>

Permissions: <http://www.sagepub.com/journalsPermissions.nav>

Citations: <http://wcx.sagepub.com/content/29/4/412.refs.html>

>> [Version of Record](#) - Oct 22, 2012

[OnlineFirst Version of Record](#) - Oct 3, 2012

[What is This?](#)

Adolescents' Disciplinary Use of Evidence, Argumentative Strategies, and Organizational Structure in Writing About Historical Controversies

Written Communication
29(4) 412-454
© 2012 SAGE Publications
Reprints and permission:
sagepub.com/journalsPermissions.nav
DOI: 10.1177/0741088312461591
<http://wcx.sagepub.com>



Susan De La Paz¹, Ralph Ferretti²,
Daniel Wissinger¹, Laura Yee¹,
and Charles MacArthur²

Abstract

This study considers how adolescents compose historical arguments, and it identifies theoretically grounded predictors of the quality of their essays. Using data from a larger study on the effects of a federally funded Teaching American History grant on student learning, we analyzed students' written responses to document-based questions at the 8th grade ($n = 44$) and the 11th ($n = 47$). We report how students use evidence (a hallmark of historical thinking), how students structure their historical arguments, and what kinds of argumentative strategies they use when writing about historical controversies. In general, better writers cite more evidence in their arguments than weaker writers, and older students demonstrate how to situate evidence in ways that are consistent with the discipline. Both the structure of students' arguments and their use of evidence were predictive of the overall quality of their essays. Finally, students' use of argumentation strategies revealed

¹University of Maryland, College Park, Maryland

²University of Delaware, Newark, Delaware

Corresponding Author:

Susan De La Paz, Department of Special Education, University of Maryland, College Park, MD 20742

Email: sdelapaz@umd.edu

patterns relevant to the historical topic and sources in question, as well as to differences related to writing skill. In our sample, better writers used strategies based on facts and evidence from the documents more so than weaker writers and demonstrated the capacity to contextualize and corroborate evidence in their arguments.

Keywords

written argumentation, adolescent literacy, history, primary source documents

The ability to generate arguments that make thoughtful contributions to historical discourse requires evaluation and interpretation of multiple sources of information, often with conflicting perspectives, in essence reflecting one's capacity for critical thinking. Despite the widely recognized importance of argumentation skills (Kuhn, 1991; van Eemeren, Grootendorst, & Henkemans, 1996), it is unclear how adolescents interpret different sources or use them to write arguments about controversial issues. On the one hand, students appear to possess requisite skills to discuss, argue, debate, and form disagreements on certain topics (Felton & Herko, 2004), and in fact, they appear capable of producing basic components of argumentation in conversation by late childhood (Anderson, Chinn, Chang, Waggoner, & Yi, 1997; Eisenberg & Garvey, 1981; Stein & Miller, 1993). On the other hand, students' written argumentation does not seem to be as well developed (Salahu-Din, Persky, & Miller, 2008). When asked to write about everyday controversies not grounded in historical contexts, only 31% of 12th-grade students' essays offered a thesis and some supporting reasons and examples (Salahu-Din et al., 2008). Furthermore, students' essays rarely acknowledge opposing positions, consider the merits of different views, or attempt to systematically respond to alternative perspectives (Ferretti, Lewis, & Andrews-Weckerly, 2009). Perhaps for these reasons, the K–12 Common Standards (Common Core State Standards Initiative, 2010) mandate that students become proficient in “logical arguments based on substantive claims, sound reasoning, and relevant evidence.”

These limitations, which are evident in students' writing about everyday controversies, presage the challenges they experience when writing disciplinary arguments. Clearly, students must differentiate everyday argumentation from argumentation in the disciplines to read and write in secondary contexts (Moje, 2008; Shanahan & Shanahan, 2008). In contrast to everyday arguments, disciplinary arguments are grounded in strategies and standards that are used by members of the disciplinary community (Ferretti & De La Paz,

2011). As students progress through the curriculum, literacy and content area learning become interrelated, making academic progress increasingly dependent on the acquisition of specialized knowledge and skills (Ferretti & De La Paz, 2011; Monte-Sano & De La Paz, 2012). In short, students are expected to argue and write like disciplinary experts (De La Paz, 2005; Ferretti & Okolo, 1996; Monte-Sano, 2008; Shanahan & Shanahan, 2008).

Stevens, Wineburg, Herrenkohl, and Bell (2005) posit that effective argumentation differs across disciplines because the epistemological criteria for judging claims are discipline specific. Historical writing shares an argumentation stance with other forms of writing. For example, the goal of science argumentation is to coordinate evidence and theory that support or refute an explanatory conclusion, model, or prediction (Suppe, 1998). Arguments about literature also ground interpretative claims in evidence drawn from the text (Lewis & Ferretti, 2011; Newell, Beach, Smith, & VanDerHeide, 2011); however, organizational decisions (e.g., use of overarching and subthemes) help the writer convey his or her interpretation (Christie & Dreyfus, 2007). Moreover, the nature of the data and the warrants—that is, the evidence and the connection between evidence and claim—is particular to the discipline (Monte-Sano, 2010; Shanahan & Shanahan, 2008). In the case of historical argumentation, the relevance of the evidence is established by warrants that link the evidence to the artifacts' sources, the perspectives of the artifacts' creators, and the historical contexts within which the artifacts were created (Hexter, 1971; Mink 1987). Given that historians privilege argumentation over other writing forms (Bain, 2006; Collingwood, 1943), it is important to explore how novices use evidence and construct arguments about controversial issues and how these skills develop in response to instruction.

Prior work suggests that without instruction, students face significant challenges when asked to write historical arguments from primary sources. Wineburg's (1991a) seminal work revealed that high school students failed to interrogate sources, primarily because they did not see the texts as written by individuals with specific purposes and intentions—even trusting textbooks as the most reliable source of information. And when writing historical arguments, students draw on source evidence indiscriminately (Britt & Aglinskias, 2002; Perfetti, Britt, & Georgi, 1995), often responding to requests to create arguments by first figuring out what they want to say, then using the documents to support their standpoint (Monte-Sano, 2008). They tend to have difficulty grasping the nature of historical context (Husbands, 1996; Shemilt, 1983), either because they lack knowledge of specific historical contexts (Halldén, 1997; Van Drie & Van Boxtel, 2008) or because they judge past actors and actions by present standards (VanSledright, 2002).

Coffin's (2006) work in linguistics highlights other challenges for students when they read and write with primary sources: recognizing historical perspectives, understanding time, identifying cause-and-effect relationships, and developing different ways of writing about the past. Schleppegrell, Achugar, and Oteiza (2004) describe how language and content are integrated and that English learners especially need support to analyze the language of the texts and the meaning implicit in that language. Fortunately, research has demonstrated that with instruction, disciplinary literacy goals are attainable for a wide range of students, including students as young as fifth grade (VanSledright, 2002), students with disabilities (De La Paz, 2005; Ferretti, MacArthur, & Okolo, 2001), and students who are English learners (Fránquiz & Salinas, 2011; Zwiers, 2006).

Wineburg's (1991b) research with historians identifies historical ways of thinking evident in experts' reading, and Monte-Sano's (2010) exploration of high school students' written historical arguments affirmed how, with disciplinary instruction, students used evidence in their writing. Five constructs emerged in these analyses: (a) factual and interpretive accuracy, (b) persuasiveness of evidence, (c) sourcing of evidence, (d) corroboration of evidence, and (e) contextualization of evidence. Factual and interpretive accuracy, persuasiveness of evidence, sourcing of evidence, corroboration of evidence, and contextualization of evidence. The way that students used evidentiary warrants demonstrated the extent of their historical thinking in their arguments. Monte-Sano found that approaching writing from a disciplinary stance required students to credibly select and situate evidence in a historical context that clarified its significance and that writing a convincing historical argument involved more than knowledge about the writing process. It also involved conceptual understanding of the historical topic, procedural understanding of historical analysis, and background content knowledge.

Writing assignments and instruction that are consistent with a disciplinary approach to history provide means to support students' historical argumentation. Young and Leinhardt (1998) found that repeated writing in response to document-based questions (DBQs) in a high school Advanced Placement U.S. history class helped students progress from listing pieces of knowledge without relating them to one another to synthesizing evidence into a unique interpretation. Writing arguments, in particular, appears to help students integrate historical content because they must interpret and organize information from historical documents in a new way (Newmann, 1990). More recently, Nokes, Dole, and Hacker (2007) demonstrated benefits in teaching students to apply Wineburg's (1991a) sourcing, contextualization, and corroboration heuristics to reading historical documents as students wrote arguments that used documents as evidence, and De La Paz and Felton (2010) found that

11th graders who learned historical inquiry and writing strategies improved the historical accuracy, persuasiveness, and quality of claims and rebuttals in their history essays. Together, these studies indicate that instruction in writing historical arguments, when combined with the analysis of historical documents, is beneficial for novices in secondary contexts.

Leinhardt's (2000) analysis of a precocious student's development in writing historical essays in response to four DBQs is particularly enlightening. These analyses focused on the organization of the student's essays, the student's use of connectors that link statements in his essays, and his use of evidence drawn from the documents. In the first essay, the student used a variety of discourse markers to link causal statements and illustrations, but he did not use the causal statements to link assertions and evidence, and the essay lacked a compelling argument. In short, the student understood some of the conventions of writing texts but did not integrate this knowledge with the substantive history content. On the fourth and final DBQ, the student was able to integrate his knowledge of writing with the history content. In comparison with his first essay, the final essay was longer, more elaborate, and more balanced across his major interpretative claims. In addition, the student used more connectors to link statements in his essay, and perhaps most important, he cited documentary sources and used them as evidence for interpretative claims. However, while sources were cited and invoked as evidence, the student failed to source the document's authors or conjecture about their motivations or relevance.

Without instruction, novices do not employ historians' standards and strategies when writing historical arguments. Unfortunately, we know little about how novices write arguments about historical issues and the strategies they use when they write about historical controversies. However, we know that in the absence of disciplinary knowledge, students often rely on prior conceptions drawn from everyday experience to reason and learn about the domain (Donovan & Bransford, 2005; Ferretti et al., 2005). For example, Brophy and Alleman (2002, 2003) showed that students often depend on their prior conceptions when reasoning about cultural universals—that is, categories of human experience that include activities related to the basic needs. For example, children were asked to explain why people might prefer to live in houses or apartments, who is paid for these accommodations and why they had to be paid for, and what distinguished renting and buying a place to live. The authors found that most children had little knowledge about these issues. Nevertheless, as in Wineburg's (1991a; 1991b) work, children reasoned from common knowledge about people's motives and goals. Rather than explaining housing arrangements in economic terms, children invoked personal and aesthetic motives to account for housing choices.

These findings suggest that novices' written arguments about historical issues may be based on their prior experience with everyday arguments. Everyday arguments are both pragmatic and dialectical; that is, they have practical aims that are achieved with interlocutors that may have different viewpoints about a controversy (van Eemeren & Grootendorst, 1992; van Eemeren et al., 1996; Walton, 1996; Walton, Reed, & Macagno, 2008). Arguments comprise a structured constellation of propositions that is meant to achieve its discursive purposes (van Eemeren & Grootendorst, 1992; van Eemeren et al., 1996). Writers' discursive purposes are accomplished by using *argumentative strategies* (Ferretti et al, 2009; Nussbaum & Edwards, 2011; Walton et al., 2008), which are conventionalized ways of representing the relationship between a standpoint and its supporting justificatory structure. The *cause-to-effect argumentation strategy*, which involves making an explicit claim that some event caused another outcome, is commonly used by historians (von Ranke, 2010) and frequently appears in social studies textbooks.

To illustrate, simple argumentation strategies include *argument from example* (i.e., a generalized proposition with illustration; is descriptive of the situation, making the case that something has happened), *argument from commitment* (you are committed to a principle or obligation; thus, you should follow through or behave in accordance with that commitment), and argument from *cause to effect* (explicit claims that some event caused an outcome). More complex strategies require logical relationships, such as argument from fear appeal:

Premise 1: If you do not bring about *A*, then *D* will occur.

Premise 2: *D* is very bad for you.

Premise 3: Therefore, you ought to prevent *D* if possible.

Premise 4: But the only way for you to prevent *D* is to bring about *A*.

Conclusion: Therefore, you ought to bring about *A*.

To date, Walton et al. (2008) have identified dozens of argument schemes that illustrate how a writer achieves his or her purpose. If novices' arguments are rooted in personal experience, then we would expect such writers to invoke these common argumentation strategies when asked to write essays in response to DBQs about controversial issues.

A recent study by Ferretti and colleagues (2009) evaluated the reasonableness of Walton and colleagues' (2008) approach, in a detailed analysis of the structure of students' written arguments on a familiar school controversy (homework). In that study, the researchers applied this approach in an examination of the extent to which structural elaboration and subordination, as well

as the kinds of strategies that elementary students used, could account for the rated quality. They also considered which specific strategies increased the reasonableness of their arguments. Structural elaboration and subordination relate to the depth of ideas that are linked in forming an argument and so provide a more nuanced account of student writing than merely counting propositions that can be categorized as claims, data, and warrants. As explained above, the latter variable examines the pragmatic function of the writer's ideas. However, while this study provided a statistical account for the variation in the quality of students' essays and qualitatively analyzed students' constellations of argument strategies, it is not possible to forecast whether this analytic approach, designed for everyday argumentation, can be extended to account for discipline-specific argumentative writing.

We know of no evidence about the efficacy of their approach for the analysis of adolescents' document-based arguments about historical controversies. In fact, there is a dearth of research about the structure of students' written arguments and the kinds of strategies they use to increase the reasonableness of their standpoints when writing about historical controversies. Consequently, a major goal of this study was to determine if Walton's argumentation theory (as applied by Ferretti and colleagues' 2009 analytic strategy) could account for the quality of students' document-based arguments about historical controversies. Second, our intent was also to explore the types of argumentation strategies that students relied on to craft arguments in response to different DBQs. Finally, we were interested in the extent to which students' use of documentary evidence (Monte-Sano, 2008; 2010) was also predictive of their essays' quality.

In sum, our analyses explore the structure of 8th- and 11th-grade students' written arguments, the kinds of argumentative strategies used by them, and how they use sources when they write document-based arguments about historical controversies. Participants were given DBQs that explicitly requested arguments about historical controversies and were provided documents that presented contrasting perspectives about these controversies. We anticipated differences that relate to the specific DBQs about which they wrote (i.e., specific events described in the sources and varying authors' perspectives on the issues). On the basis of Ferretti and colleagues' (2009) work, we also anticipated that better writers would develop more elaborate arguments that evidenced a greater degree of structural subordination. Furthermore, we expected that better writers would refer to more documents and show a more sophisticated use of these documents than less able writers. Finally, as in other studies of writing competence, we anticipated that grade level would also be predictive of performance (Graham & Perin, 2007).

Research Questions

Research Question 1: How do 8th- and 11th-grade novices structure their essays when writing in response to DBQs about historical controversies? Do novices use everyday argumentative strategies when writing their arguments? If so what argumentative strategies do they use?

Research Question 2: What patterns in students' written responses are common between 8th- and 11th-grade students as they attempt to use historical evidence from documents in their writing? What patterns are common between students who are strong writers in comparison to those who are weak writers as they attempt to use historical evidence from documents in their writing?

Methods

Participants and Setting

Participants came from a larger study (De La Paz, Malkus, Monte-Sano, & Montanaro, 2011), which included 8th- and 11th-grade teachers in northern California who participated to varying degrees in a Teaching American History professional development program aimed at promoting historical thinking and writing skills in students. The present study is a retrospective analysis of selected students' work from that project, and it includes any student who met specific criteria (see below), regardless of his or her teacher's degree of involvement in the Teaching American History program. Because teachers in the original study participated in a yearlong intervention, we decided to restrict our sample to student writing at the beginning of the year, to avoid comparing students' compositions from teachers with unequal levels of participation in professional development. In the original study, there were five cooperating school districts with varying socioeconomic levels, from urban to suburban settings.

Materials and Writing Task

Approaching history as an inquiry into the past that fosters analyzing evidence, developing arguments, and conveying these interpretations in writing supports students' learning in the discipline (Bain, 2005; Holt, 1990). Therefore, we asked students to compose essays in response to DBQs, each

with four or more document excerpts (“document sets”) about controversies in American history. Our format for prompting students to write responses to primary sources was similar to the way that the New York State exam is framed. As such, we provided students with a brief historical context and document excerpts and prompted them to use “specific details from at least three documents” in an introduction, supporting paragraphs, and conclusion that “responded to the (historical) question.”

Content for the historical questions was based on state and district standards, and individuals involved in the larger study from which the current data come helped develop DBQs that were aligned with the cooperating district’s calendar of instructional objectives for the year. Each document set provided opposing positions and contrasting information on a single topic. The 8th graders responded to a question about the Mexican-American War, “*Did the U.S. government have a reasonable (or unreasonable) argument for going to war with Mexico?*” The 11th graders responded to a question about the Progressive Era, “*Who had the better vision for improving the conditions of African-Americans during the early 1900s, Booker T. Washington or W.E.B. DuBois?*” Students were asked to evaluate the issues and take a position in a written argument, using the documents for support.

A university historian, district librarians, and the first author developed the primary source content collaboratively, by creating document excerpts appropriate for each grade level. Moreover, in keeping with recommendations by Wineburg and Martin (2009), we adapted the primary sources by italicizing difficult words and presenting synonyms in square brackets (e.g., “*orator* [speaker]”) to help students with unfamiliar vocabulary. Finally, we asked teachers to assign the DBQ prompts before students learned the corresponding topic in school because we were unable to standardize teachers’ presentation of the content.

Screening process for identifying participants. We developed a two-stage process for selecting participants based on our overarching purposes for the study. One of our goals was to compare students at each grade level, which we did by virtue of having 8th- and 11th-grade students. We also were interested in comparing how students with different levels of writing proficiency approached the DBQ task, but we did not have standardized writing test scores for students in the original project. Therefore, we developed a set of criteria to categorize students into two different writing ability groups, “good” and “poor” writers, using qualitative scores from their performance on the pretest DBQ, because it was given before teachers began to participate in our professional development.

Writing ability. We identified good versus poor writers using a 7-point primary trait rubric adapted for argumentative writing (scores of 0-6; Ferretti,

MacArthur, & Dowdy, 2000) and used to assess overall quality. This rubric was based on prior work by Ferretti et al. (2000) and was developed to gauge the writer's ability to (a) provide a clear opinion on the topic; (b) support a position with accurate facts, examples, and details; (c) weigh the importance, reliability, and validity of the evidence; (d) analyze conflicting perspectives presented in the documents; and (e) include a strong introduction and conclusion. Undergraduate students majoring in history, as well as a practicing social studies teacher earning his master's degree in education, independently scored all essays in the original data set. All papers were read independently by two individuals, with subsequent interrater agreement (within 1 point) at 90% for the 8th-grade level and 95% for the 11th-grade level.

We used pre- and posttest¹ DBQ writing samples collected in the larger study mentioned above to define groups for the current study but subsequently used only the pretest data for the present purpose of exploring aspects of argumentative writing. We defined good writers in the 8th and 11th grades as performing above the mean (score of 2.5 and 3.0, respectively) on both assessments. At the 8th-grade level, poor writers scored at or below 1.0 on both assessments, and the 11th-grade poor writers scored at or below 2.0 on both assessments. Thus, we required at least a 1-point difference in trait score between good and poor writers at each grade level to distinguish between good and poor writers. Prior research (cf., Graham & Perin, 2007) has suggested that a single-point difference in quality can meaningfully differentiate students' writing ability.

Reading comprehension. Because we knew very little about students in the original study in terms of their background learning characteristics, we took additional steps to ensure that we were selecting students who were able to read and demonstrate adequate literal comprehension of the primary sources in the DBQ assessment before being asked to write a historical argument. Therefore, in addition to selecting students on the basis of their qualitative writing score, we examined the quality of students' responses to open-ended factual questions written for each document. We did this because poor literal comprehension of the documents would severely limit students' capacity to write a coherent historical argument, and we wanted to ensure that students who were identified as poor writers were not confused with students who did not understand the task materials.

For example, on the 11th-grade question about the Progressive Era, with respect to an excerpt from Booker T. Washington's 1901 book *Up From Slavery*, students were asked to answer the following factual question: "What did Booker T. Washington believe African Americans must do in order to improve their social position?" Responses to this question and the remaining

comprehension questions were open-ended, as students could use any of several ideas from the document to support their answer. Moreover, because responses ranged in overall quality, we graded all responses on a 5-point scale (0 = *no understanding*, 5 = *complete understanding*). We limited the selection of participants to students whose average comprehension score was 3.0 or better (across the full document set), establishing that students essentially needed to provide reasonable answers to at least three of the five documents.

While it may be argued that this criterion may not be sufficient for demonstrating full comprehension, we felt it was appropriate for two reasons. First, the DBQ task directions prioritized writing a historical argument rather than answering questions about each primary source. Perhaps because of this, the majority of students wrote brief answers, and many skipped a response to the final document, either to save time or because they may not have found the need to respond to the last question. Thus, our criterion for average comprehension score was reasonable given the DBQ task. Second, the DBQ document sets were written with contrasting perspectives (a minimum of two documents for each perspective). Because of this, students could have responded to the historical question based on reading some but not all primary sources. We believed that it would be unfair, then, to require that students demonstrate their comprehension of the full document set to be in the current study. When students' answers were scored for comprehension, all of the students' responses were rated independently by two readers, with the resulting percentage agreement for adjacent scores (within 1 point) being .80 for 8th-grade students and .99 for the 11th-grade students.

Final sample. Based on this two-part screening process, our final sample included 44 eighth graders and 47 eleventh graders (see Table 1 for descriptive information). There were 28 good and 16 poor 8th-grade writers as well as 33 good and 14 poor 11th-grade writers. More of the good writers were girls: 46% of the good 8th-grade writers and 58% of the good 11th-grade writers. Very few students received services for special education needs (1 or 2 poor writers at each grade level). Our sample was ethnically diverse—26% White, 22% African American, 24% Asian, 21% Hispanic, 5% Vietnamese, and 1% Filipino—and representative of the participating school districts. However, our samples of good writers were overrepresented by African American students at the 8th grade (57% of the good 8th-grade writers were African American) and overrepresented by Asian American students at the 11th grade (61% of the good 11th-grade writers were Asian American).

We also wished to establish the degree to which students were proficient in written English, so we examined two sources of information to determine their abilities. The school districts provided information about students' English

Table 1. Participant Characteristics, *n*

	Eighth Grade	Eleventh Grade
Total		
Good writers	28	33
Poor writers	16	14
Female		
Good writers	13	19
Poor writers	4	6
Receiving special education		
Good writers	0	0
Poor writers	1	2
White		
Good writers	8	8
Poor writers	5	3
African American		
Good writers	16	0
Poor writers	2	2
Asian		
Good writers	0	20
Poor writers	1	1
Hispanic		
Good writers	1	4
Poor writers	6	8
Vietnamese		
Good writers	3	0
Poor writers	2	0
Filipino		
Good writers	0	1
Poor writers	0	0
English language arts ^a		
Good writers	435.3 (42.5)	386.9 (44.6)
Poor writers	343.8 (49.5)	327.9 (42.9)

^aMean score (SD).

proficiency status for about 60% of the good and poor 8th-grade writers, and we found that 38% of the good writers and 21% of the poor writers spoke a language other than English before beginning their academic careers (the districts did not provide information on level of English proficiency, only whether students were considered “English only” or “not English only”). With respect

to the 11th-grade students, we had information about their English proficiency for 80% of the good writers and found that 49% initially spoke a language other than English. Fewer than 10% of the poor 11th-grade writers were missing information regarding English proficiency, and within this subgroup, 32% had been identified as speaking a language other than English when beginning school. Therefore, while the cooperating districts did not share information about students who were not native English speakers' relative proficiency in English, this information shows that individuals who were considered English learners were included in our good and poor writer groups, with more English learners among the former than the latter.

Our designation for considering students as good and poor writers appears to have additional face validity in terms of their performance on the English-language arts portion of the California Standards Test (Educational Testing Service, 2004), which was mandated by the state to measure students' progress toward achieving state-adopted content standards for each grade. The cooperating school districts provided students' scaled scores, which revealed that the good writers at each grade level met state criteria as being proficient in language arts (average scores met or exceeded a benchmark criterion of 350) whereas poor writers at each grade level were on average considered to be functioning at basic levels of language arts proficiency. It is important to note that poor writers did not perform at the state's lowest level, below basic.

A one-way analysis of variance that compared good and poor writers on English-language arts scores showed significant differences, $F(1, 87) = 24.08$, $MSE = 1,988.389$, $p = .000$. Five pairwise comparisons among the means for writers of different ability were significant, controlling for Type I error across the three tests at the .05 level by using Holm's sequential Bonferroni procedure. These results indicated that English-language arts scores were different for all groups with one exception—the poor writers at the 8th- and 11th-grade grade levels did not differ from each other. Moreover, good 8th-grade writers on average met the California state criterion as advanced performers on the English-language arts test, whereas good 11th-grade writers on average met state criterion as proficient performers. Finally, the scores of the poor 8th- and 11th-grade writers ranged from basic to proficient on the English-language arts test (Educational Testing Service 2004).

Dependent Measures

Evidence. Three variables were developed to capture the extent to which participants interpreted documentary evidence in their essays: (1) we counted the number of document citations—quotes and paraphrases from quotations

that were clearly taken from the documents that students read; (2) we noted how many different (unique) documents students used in their essays from the set of four or more; and (3) we explored one specific student learning goal related to disciplinary thinking and literacy—namely, how students were using evidence from documents in their written arguments. We further privileged students' use of quotations and presence of sourcing over their use of facts and examples, because these forms of evidence provide an indication of the students' recognition of an author's perspective and because examining evidence in this way allowed us to better understand relationships between this analysis and our subsequent analysis related to the ways that students crafted their arguments. A 6-point rubric was developed (based on De La Paz & Felton, 2010; Monte-Sano, 2010) indicating the relative sophistication of document use (henceforth, *highest level of document use*), from 0 (does not refer to documents) to 5 (evaluates the quote or evidence or uses it as a means to further his or her argument; see Table 2).

Each element of the rubric was written to capture difference in students' historical thinking. A history educator who was skilled in the analysis of students' historical thinking confirmed the instrument's content validity. The first author scored all essays. Interrater reliability was calculated by the first author and an independent reader who was unfamiliar with the design and purpose of the study, using a random sample of 25% of each type of writer across topics, with 100% agreement for the number of documents and 100% agreement for the number of unique documents. Interrater reliability results for the highest level of document use was as follows: 86% for good 8th-grade writers, 100% for poor 8th-grade writers, 86% for good 11th-grade writers, and 100% for poor 11th-grade writers at pre- and posttest.

Argumentative structures. The argumentative structure of each essay was depicted through a detailed graphing process. Each essay was graphed separately, and the components depicted in the graph were then examined to establish the overall structure. This approach to analysis, based on *pragma-dialectical theory* of argumentation (van Eemeren & Grootendorst, 1992, 2004; van Eemeren, Grootendorst, & Henkemans, 2002), identifies the stance taken by the writer, as well as alternative standpoints identified by the writer, along with coordinated and subordinated supports for these standpoints. To clarify, an important difference between the two types of reasons are that in subordination, each succeeding reason is a layer in the argument that buttresses the preceding reason. Additional argumentative elements identified in the graphing process were rhetorical, counterargument, rebuttal, nonfunctional, and conclusive statements. The resulting graphic representation depicted the elements of the argumentative structure and the superordinate and subordinate relationships among these elements in each participant's written argument.

Table 2. Use of Evidence

Evidence Level	Example
0. The writer does not refer to documents in the essay.	—
1. The writer mentions a specific document or a specific author or quotes from a source—but the evidence is misunderstood or inaccurate.	I think the US had a good reason to war with Mexico for many reasons president Polk said and other stuff he did but [couldn't say].
2. The writer mentions a specific document or a specific author or quotes from a source—and the evidence is understood or accurate or contextually relevant.	“Measured by any standard, white of black, Washington must be regarded today as one of the great men of this country; and in the future he will be so honored.” We can see today that this is true.
3. The writer explains or interprets a quote or uses a quote to substantiate a claim. (The citation may not be substantive, however.)	Abraham Lincoln said that the Americans shouldn't have went (sic) to war with the Mexicans. He believed that the Americans were moving into hostile territory.
4. Meets criteria for a 3, as well as the following: selection of the evidence is balanced (e.g., there are two quotes with more than one perspective) or significant (others who read the document would find the quote to represent a substantive idea or perspective).	Thirdly, I don't believe Swain when he says to “redeem the Mexican people from . . . tyranny and to facilitate the entire removal of those rivals from this continent.” By doing that, the United States is stealing territory and “violated the sovereignty of nations.”
5. Meets criteria for a 4 as well as the following: evaluates a quote or uses the quote as a means to further an argument.	During the early 1900s, Booker T. Washington had a better vision for improving the conditions of African Americans. . . . “The whole future of the Negro rested largely upon the question as to whether or not he should make himself, through his skill, intelligence, and character, of such . . . value to the community in which he lived that the community could not dispense [do without] with his presence.” Washington is saying that a person's skill and intelligence changed their conditions if he could be of much value to the community. Washington's theory was that as long as a person learned to do something better than someone else he would be respected. Hard work was also something that African Americans should do according to Washington to improve conditions.

Ferretti and colleagues' (2009) process for identifying essay structure was used to graph each essay to determine its structural elements and level of elaboration and subordination in each essay (see Table 3 for criteria used to identify each

Table 3. Structural Elements for Sample of Good 11th-Grade Essay

Criteria	Description	Example
Introduction (96%) ^a	Content that foreshadows the student's argument	Work, work, work. Workers were all that slaves were looked at as, and even then some were punished because they didn't do things right. From many years of being racially superior to blacks, white people were not going to automatically be equal with them after the Emancipation Proclamation. They were still going to be looked at as nothing people who should do everyone else's work. Two main people who wanted to improve conditions for African Americans were Booker T. Washington and W.E.B. Du Bois. Both had different outlooks on how to achieve improvements, but I think that Du Bois had the best goals and visions.
Standpoint (100%)	The premise advanced by the student	Du Bois had the best goals and visions.
Level 1 reasons for the author's standpoint (88%)	Propositions offered as direct support for the premise	From duBois' point of view rights should be given to African Americans first
Level 2 reasons for the author's standpoints (88%)	Propositions offered as direct support for the reasons	e.g., should first be able to vote
Level 3 reasons for author's standpoints (86%)	Propositions that support Level 2 reasons	e.g., basically said that blacks should have the same rights as whites
Level 4 reasons for the author's standpoints (93%)	Propositions that support Level 3 reasons	—
Counterarguments (88%)	Potential criticisms of either the student's standpoint or reasons for the student's standpoint that could be used to bolster the alternate standpoint	e.g., main idea for conditions to become better would be for African Americans to earn their way into society.
Rebuttals of the alternative standpoint (87%)	Propositions that either attack an alternate standpoint or undermine counterarguments	e.g., think that this is totally opposite in what ex-slaves wanted.

(continued)

Table 3. (continued)

Criteria	Description	Example
Conclusion (100%)	Summary of what the student has written	Between Washington and Du Bois, I think that Du Bois had the better ideas for the people themselves. He wanted them to have equal rights and to not be discriminated against. Washington basically said that they weren't equal and that they would have to earn their rights. Although it would take awhile for relations to become better, everybody should be equal from the beginning.

Note: See textbox for essay.

*Percentage of agreement for a random set of 25% of papers for each grade and ability level.

Work, work, work. Workers were all that slaves were looked at as, and even then some were punished because they didn't do things right. From many years of being racially superior to blacks, white people were not going to automatically be equal with them after the Emancipation Proclamation. They were still going to be looked at as nothing people who should do everyone else's work. Two main people who wanted to improve conditions for African Americans were Booker T. Washington and W.E.B. Du Bois. Both had different outlooks on how to achieve improvements, but I think that Du Bois had the best goals and visions.

Washington's main idea for conditions to become better would be for African Americans to earn their way into society. They would have "to make himself, through his skill, intelligence, and character . . . value to the community." This at least states that blacks do have character, skill, and intelligence, but it also says how they still have to push their way into society. Washington also said that in order to earn respect an African American should learn "to produce what other people wanted and must have." I think that this is totally opposite in what ex-slaves wanted. They didn't want to have to work for their respect, they wanted to already have it and to fit into society as easily as possible. Since white people weren't going to give them respect as quickly and easily, why should they try hard to earn it?

From du Bois' point of view, certain rights should be given to African Americans first. They should first be able to vote, and then along with that will come. . . . "Freedom, manhood, the right to work, and the chance to rise . . ." Du Bois also wanted discrimination to stop, laws to be enforced equally among different classes and races, and for children to be educated well. He believed that African Americans should be freemen " . . . to walk, talk, and he with [those who] with to be with us . . ." I go along with everything that Du Bois wanted. He basically said that blacks should have the same rights as whites, and with that maybe things will improve. I think it's better to start out with basic equality and work from there instead of starting with working to gain basic equality.

Although I disagree with Washington, he had those who supported him and also those who disagreed with him, as well. One who supported him was a journalist and an author, Ray Stannard Baker. He said that whenever he saw good things, like a thriving business, a good home, or friendly relations between whites and blacks, he would think of Washington. But who's to say that Du Bois didn't do this, or natural occurrences made things better on their own? A person who also disagreed with Washington was Ida B. Wells, a black social worker and journalist. She believed that he steered away as far as he could from antagonism, that he didn't join them because they were too extreme, and that he shouldn't be there advisor if he didn't have political strength in his own race.

Between Washington and Du Bois, I think that Du Bois had the better ideas for the people themselves. He wanted them to have equal rights and to not be discriminated against. Washington basically said that they weren't equal and that they would have to earn their rights. Although it would take awhile for relations to become better, everybody should be equal from the beginning.

element of the organizational structure, with a description and an example that correspond to a sample essay; see textbox for essay). This process facilitated our aim in analyzing and evaluating the operative relationships among elements of argumentative discourse. The content in Table 3 is from one 11th-grade good writer. Moreover, the current example provides evidence of a more sophisticated degree of structural subordination because it contains all of the elements in one essay. Papers that were less elaborated typically contained fewer levels of subordination or advanced a counterargument without a corresponding rebuttal.

Moreover, the corresponding structure for this essay, shown in Figure 1, reveals that the student offered a simple introduction that answered the historical question. She offered one standpoint, which had one supporting Level 1 reason with counterarguments preceding and following. Each counterargument and each reason were elaborated by a series of coordinated Level 2 reasons and had one or two additional levels of subordination. The student advanced her argument with three rebuttals and offered one rhetorical statement. She cited three documents (two of them twice). Finally, the writer concluded her essay with a summary of her standpoint.

The third author independently graphed the structure of all essays. He then created a scoring guide modeled after Ferretti and colleagues' (2009) suggestions that explained how to graph the argumentative structure of each essay, and he worked with an independent rater unfamiliar with the design and purpose of the study to establish interrater reliability. Four 1-hour sessions were used for training, which included reliability checks on 15 sample papers taken from the data set. Interrater agreement (exact agreement) was computed for

each structural element (see Table 3) on the remaining papers using a randomly selected pool (25% of each type of writer and topic). Interrater agreement (agreement = agreements/agreements + disagreements) was computed for each argumentative element on a randomly selected pool (25% of the papers for each type of writer) with the following results: authors' standpoint(s) = 100%, Level 1 reasons for authors' standpoints = 88%, Level 2 reasons for authors' standpoints = 88%, Level 3 reasons for authors' standpoints = 86%, Level 4 and below reasons for authors' standpoints = 93%, introductions = 96%, conclusions = 100%, counterarguments = 88%, rebuttals = 87%.

Argumentation strategies. Based on the graphical depictions of participants' arguments, each essay was then analyzed to identify the type of argumentation strategy used for each node. The first author identified and categorized sentences and phrases from all students' essays into each strategy present in a given paper, using Walton's criteria (see Table 4 for elements, criteria, and examples). The fourth author then identified and categorized sentences and phrases from a randomly selected pool of 25% papers from each type of writer and topic. They identified a total of 22 different strategies. Interrater reliability for each type of strategy (exact agreement) in each paper for the Mexican-American War topic was 82% for good writers and 80% for poor writers. Reliability for each type of strategy for the Progressive Era topic was 93% for good writers and 90% for poor writers.

Length. Total words were counted with a word processing program (Microsoft Word 2004). We typed students' handwritten essays but did not correct spelling or grammatical errors.

Results

Our rationale in analyzing students' written arguments using this approach was multifaceted. We focused on students' use of evidence as a means to discern their disciplinary thinking. Moreover, we wished to explore the underlying discursive purposes of students' arguments because the topics under investigation involved historical controversies, which are commonly considered in secondary social studies curriculum, and have previously not been explored. Finally, these analyses were conducted as compared to two traditional measures of students' writing: the length of their papers and the depth and range of their elaborations.

Evidence

A series of 2 (ability) \times 2 (grade) analyses of variance were used to evaluate the relationship between type of writers and each dependent measure: (1) total

Table 4. Examples of Argument Strategies

Category	Description	Example
Rule	Everyone in a particular group acts or must act in a particular way because it is obligatory according to the rule.	The United States violated the sovereignty of nations. [or] They had no right to take over the land, which was owned by Mexico.
Verbal classification	Intuit a property (often ambiguous) about person, place, situation; then used as a basis for the argument.	It also seems like the Mexican people don't need saving from tyranny. [or] DuBois' vision is more of a risk, and volatile because you don't know what would go on in the process of protesting. DuBois is also just a speaker for the African Americans. He "is not a leader of mean, as Washington is. He is rather a promulgator:"
Example	Generalized proposition with illustration; making the case of something that has happened; can begin with specific markers—"For instance ..." or "Here's a case ..."	W.E.B. Du Bois knew what needed to be done and what was most important. There were about six things that needed to change. The first thing to change is the right to vote. African Americans should be allowed to vote. Second, discrimination needs to stop. Third, they are to be known as freemen opposed to slaves. Forth, the laws must be enforced both ways. The rich as well as the poor. Fifth, the children needed to be educated. They need to wipe out illiteracy. Sixth, education for real men. A college education. Not on how to work, but how to read and write.
Consequence	Policy issues (slavery, taxation) put into effect, and as a result, there were consequences (negative or positive); can begin with specific syntax, such as "if-then" statements.	By ending discrimination, they would be able to live their life with less fear from someone trying to hurt them because of their race. [or] If America did not stand up to Mexico to defend her territory, there would have been many disagreements.

(continued)

Table 4. (continued)

Category	Description	Example
Commitment	<p>“You committed to a principle or obligation, you should follow through or behave in accordance with that commitment.”</p> <p>“You didn’t follow through, so I don’t have any obligation to follow through.”</p>	<p>The Americans were just doing their job. [or] We had claimed the area and Mexico knew that.</p>
<p>Inconsistent commitment</p> <p>Expert opinion</p>	<p>“You didn’t follow through, so I don’t have any obligation to follow through.”</p> <p>Expert in domain is in a special position to know about propositions in that domain; expert beliefs as reason for why action was taken.</p> <p>Generally accepted as true.</p>	<p>Mexico also failed to keep its promise to pay millions of dollars to American citizens, which justifies the U.S. invading Mexico. Lincoln believes different and said if you can prove that the first bloodshed was on our soil maybe I’ll agree . . .</p>
<p>Popular opinion</p> <p>Slippery slope</p>	<p>Generally accepted as true.</p> <p>“This happens and as a result this happens and as a result . . .” Only mark when explicitly causal—usually improbable connections unlikely or difficult to establish; otherwise, mark as argument from consequence.</p>	<p>People witnessed first hand the damage made on their country. [or] That means people started to disagree with President Polk. I could declare that I own all the land I can see from my house, but that wouldn’t make it mine. [or] Personally, Mr. Stimson just didn’t want his or the country’s ego being hurt. If the US didn’t give a warning because of that reason, then our country sacrificed many innocent lives in sake of an ego.</p>
Sign	<p>Finding or observation X is taken as evidence of the existence of a property or event E in a given situation.</p>	<p>There is a huge gap between the blacks and white in the 1900s [sic] only because the difference between the black and white is made to be so visible. [or] It was also wartime so the use of force was expected and he who had the best weapons was expected to use them.</p>
Authority	<p>When the argument is due to authority—“I am the parent, it’s my house.”</p>	<p>If they were being cheated they could of talked to the President and [forced] him to explain everything.</p>

(continued)

Table 4. (continued)

Category	Description	Example
Cause to effect	Making an explicit claim that some event caused another outcome.	The American's [sic] invaded so Mexico had no choice but to fight. [or] By making Texas part of America it would take away the people's independence and freedom.
Fear	<p>Premise 1: If you do not bring about A, then D will occur. Premise 2: D is very bad for you.</p> <p>Premise 3: Therefore, you ought to prevent D if possible. Premise 4: But the only way for you to prevent D is to bring about A.</p> <p>Conclusion: Therefore, you ought to bring about A.</p>	<p>America was right to go to war, not just for the land, but for safety. [or] It is better to conquer Mexico sooner than waiting for it to get conquered by an enemy.</p>
Danger	<p>Premise 1: If you (the respondent) bring about A, then B will occur. Premise 2: B is a danger to you. Conclusion: Therefore (on balance) you should not bring about A.</p>	<p>That hostile take over and the unjust spoliation almost endangered the peace of the world. [or] Mexico's President thought that the U.S. was at fault because the U.S. wanted to take over territory that wasn't their and [they] endangered the peace of the world.</p>
Precedent	<p>Case-based reasoning where citing particular case is used to argue for changing an existing rule or adding a new rule to supplement existing rules.</p>	<p>Once a black man establishes his role in socioeconomic terms ... he himself will be considered equal to any white men.</p>
Perception	<p>Person P has an image of a perceptible property. To have that image is the prime reason to believe that the circumstances exemplify that image.</p>	<p>Expanding the borders of Texas was dangerous because if Mexico hadn't consented, the U.S. would not only have a bad impression, but other nations might think that the U.S. would try to take away other nations' independence and begin to manipulate them.</p>

(continued)

Table 4. (continued)

Category	Description	Example
Moral justification	Everybody in group G does X. Doing X shows that X is an acceptable norm of conduct for group G. I am a member of group G. Therefore, my doing X is morally justified as an acceptable action or policy.	A large amount of Mexican troops massacred American troops. ... Mexico had repeated threats to the USA. America has a right to defend her country and people.
Analogy	An argument from one case that is said to be familiar with another.	This is a little like the same thing as if Mexico invaded Georgia, claiming it was their land.
Position to know	One party has reason to presume that another party has access to information or knowledge that the first party does not have direct access to.	Booker T. Washington was a leader. He brought many into organization and gave many strength and the courage to move on. As Ray Stannard Baker says, "He has given them a working plan of life."
Agent from past action	An observed event appears to have been brought about by some agent A. The bringing about of this event fits a certain character quality Q. Agent A has Q. Agent A brought about the event in question.	DuBois did help launch the national Association for the Advancement of Colored People (NAACP) which was an organization that fought against racial discrimination not only just to blacks but to all minorities. DuBois didn't ask the government for a change, but instead demanded it. He fought the system for the people and didn't try to find another "easier" way to succeed.
Values	A claim of values that the author identifies as from another individual rather than a value that the author ascribes.	Washington called for the development of a skillful, intelligence, diligent black community. [or] DuBois wanted the discrimination between Black and whites to stop.
Nonfunctional	Content that is irrelevant to the topic.	People probably didn't like them as much. [or] Instead of living low, Washington believed one should take life as normal. Show off rather than hide. Wear your skin color over your clothes.

Table 5. Length, Persuasive Quality, and Document Use

Condition: Grade	Good Writers		Poor Writers	
	M	SD	M	SD
Length				
8th	219.32	61.74	76.50	27.41
11th	443.18	88.84	158.00	62.32
Total number of citations				
8th	1.32	1.81	0.44	0.63
11th	3.61	2.93	1.00	1.30
Number of unique documents				
8th	0.79	0.79	0.37	0.50
11th	2.12	0.93	0.79	0.98
Highest level of document use				
8th	1.68	1.68	0.56	0.82
11th	4.00	1.46	1.00	1.57

number of documents used, (2) number of unique documents, and (3) highest level of document use. Table 5 presents descriptive information for each measure.

Statistical analyses for the total number of documents used showed main effects for grade, $F(1, 87) = 9.05, p = .003$, and ability level, $F(3, 87) = 13.59, MSE = 4.48, p = .000$. The interaction between ability level and grade was not significant, $F(1, 87) = 3.31, p = .07$. These results indicated that older writers used more documents than younger writers and that better writers used documents more than poor writers. In addition, good 8th-grade writers used more documents than poor 11th graders.

Statistical analyses for the number of unique documents used showed main effects for grade, $F(1, 87) = 21.99, p = .000$, and ability level, $F(3, 87) = 13.88, p = .000$. The interaction between ability group and grade was also significant, $F(1, 87) = 6.17, MSE = 0.694, p = .015$. An examination of group means shows that older, better writers used more unique documents in their papers than younger, weaker writers. In addition, good 8th-grade writers used essentially the same number of unique documents as poor 11th-grade writers.

Statistical analyses for the highest level of document use showed main effects for grade, $F(1, 87) = 20.86, p = .000$, and ability level, $F(3, 87) = 44.48, p = .000$. The interaction between ability group and grade was also significant, $F(1, 87) = 6.59, MSE = 2.12, p = .012$. These results indicated that

older, better writers had the most sophisticated level of document use and that younger, weaker writers showed the least developed use of documents. Moreover, poor 8th graders typically failed to use documents at all, and good 8th-grade writers typically mentioned a specific document or a specific author or quoted from a source; however, they presented evidence fully about only half the time (mean score for highest level of document use was 1.68), and half the time, their use of evidence was not contextually relevant or was inaccurate. Table 2 provides an example: “I think the US had a good reason to war with Mexico for many reasons president Polk said and other stuff he did but [couldn’t say].” Older poor writers were similar in their use of documents as the good younger writers. Finally, the good 11th-grade writers tended to select evidence that was balanced (e.g., showing quotes with more than one perspective):

Thirdly, I don’t believe Swain when he says to “redeem the Mexican people from . . . tyranny and to facilitate the entire removal of those rivals from this continent.” By doing that, the United States is stealing territory and “violated the sovereignty of nations.”

Good 11th-grade writers also tended to select evidence that was significant (i.e., representative of a substantive idea):

During the early 1900s, Booker T. Washington had a better vision for improving the conditions of African Americans. . . . “The whole future of the Negro rested largely upon the question as to whether or not he should make himself, through his skill, intelligence, and character, of such . . . value to the community in which he lived that the community could not dispense [do without] with his presence.” Washington is saying that a person’s skill and intelligence changed their conditions if he could be of much value to the community. Washington’s theory was that as long as a person learned to do something better than someone else he would be respected. Hard work was also something that African Americans should do according to Washington to improve conditions.

This example is typical, as older and better writers frequently used evidence to substantiate a claim.

Argumentative Structures

We performed a block-entry hierarchical regression to examine the effect of the structural elements and students’ writing ability (good versus poor) on

predictions of overall quality for each topic. Theoretical predictions and prior research (Ferretti et al. 2009) led us to expect that the structural aspects of students' arguments (i.e., the depth of students' supporting justifications, consideration of alternative perspectives, and efforts to rebut alternative perspectives) would predict the essays' primary trait rating. Prior analyses of students' historical writing (De La Paz, 2005; De La Paz & Felton, 2010; Monte-Sano, 2010) led us to expect that students' use of documents would contribute to the prediction of the overall quality of students' papers. The presence of additional elements, such as introductions (which foreshadow the argument) and conclusions (which summarize it), were also used to predict the essays' persuasiveness. Finally, we expected that information about students' overall quality score on the DBQ would account for less variance in persuasiveness after the structural aspects of students' written arguments and historical thinking were accounted for. The reason is that the elaborateness of their arguments, which is captured in the structural analyses, and their use of historical documentation are directly related to the essays' quality (see Ferretti et al., 2009). Therefore, blocks were entered in the following order: "my side" (author's standpoints, Level 1 reasons, Level 2 reasons, Level 3 reasons, and reasons at or below Level 4), "document use" (number of documents, unique documents, and most sophisticated use of documents), "extras" (introduction and conclusion), and "writing ability" (good vs. poor writer).

Mexican-American War. Table 6 presents standardized regression coefficients (β) and semipartial coefficients (sr_i) for independent variables at each step of the analysis: after Step 1, with only "my side" elements included in the regression equation, $R^2 = .52$, $F(4, 39) = 10.51$, $p = .000$; after Step 2, with the addition of "document use" elements in the equation, $R^2 = .74$, $F(3, 36) = 9.91$, $p = .000$, indicating that the extent to which students were able to cite and think about documents also affected the overall ratings of their essays. After Step 3, R^2 increased to $.80$, $F(2, 34) = 5.36$, $p = .010$, indicating that the students' introductions and conclusions added to the predictive power of the essay rating. Finally, after Step 4, R^2 increased to $.94$, $F(3, 33) = 74.87$, $p = .000$, indicating that writing ability also contributed significantly to the overall prediction among the eighth graders.

Progressive Era. Table 6 presents standardized regression coefficients (β) and semipartial coefficients (sr_i) for independent variables at each step of the analysis: after Step 1, with only "my side" elements included in the regression equation, $R^2 = .44$, $F(5, 41) = 6.52$, $p = .000$; after Step 2, with the addition of "document use" elements in the equation, $R^2 = .73$, $F(3, 38) = 13.26$, $p = .000$, indicating that the students' document use again reliably affected the overall persuasiveness of their essays. After Step 3, R^2 increased to $.77$, $F(2, 36) =$

Table 6. Block-Entry Hierarchical Regression Analysis for Prediction of Persuasiveness: Mexican-American War

Variable	β	sr_i
Step 1		
Intercept		
Level 1 reasons	.40	.49
Level 2 reasons	.39	.44
Level 3 reasons	.23	.29
Level 4 or below reasons	.12	.17
Step 2		
Intercept		
Level 1 reasons	.39	.58
Level 2 reasons	.37	.53
Level 3 reasons	.19	.31
Level 4 or below reasons	.18	.32
Number of documents	.08	.10
Number of unique documents	.30	.26
Highest rating of document use	.13	.15
Step 3		
Intercept		
Level 1 reasons	.28	.46
Level 2 reasons	.30	.47
Level 3 reasons	.16	.31
Level 4 or below reasons	.14	.28
Number of documents	.03	.04
Number of unique documents	.23	.23
Highest rating of document use	.16	.21
Introduction	.31	.45
Conclusion	-.01	-.02
Step 4		
Intercept		
Level 1 reasons	.02	.05
Level 2 reasons	.11	.30
Level 3 reasons	.04	.14
Level 4 or below reasons	.13	.43
Number of documents	-.10	-.23
Number of unique documents	.23	.39
Highest rating of document use	.13	.29
Introduction	.09	.25
Conclusion	-.08	-.21
Type of writer	-.73	-.83

(continued)

Table 6. (continued)

Variable	β	sr_i
Step 1		
Intercept		
Level 1 reasons	.34	.40
Level 2 reasons	.29	.34
Level 3 reasons	.39	.39
Level 4 or below reasons	.08	.09
Standpoint	.13	.16
Step 2		
Intercept		
Level 1 reasons	.17	.35
Level 2 reasons	.21	.34
Level 3 reasons	.20	.34
Level 4 or below reasons	.07	.12
Standpoint	.04	.08
Number of documents	.08	.13
Number of unique documents	.11	.13
Highest rating of document use	.47	.53
Step 3		
Intercept		
Level 1 reasons	.11	.19
Level 2 reasons	.16	.29
Level 3 reasons	.17	.29
Level 4 or below reasons	.11	.19
Standpoint	-.00	-.00
Number of documents	.05	.08
Number of unique documents	.15	.18
Highest rating of document use	.38	.45
Introduction	.09	.12
Conclusion	.20	.29
Step 4		
Intercept		
Level 1 reasons	-.07	-.17
Level 2 reasons	-.02	-.05
Level 3 reasons	.02	.06
Level 4 or below reasons	.00	.01
Standpoint	.08	.22
Number of documents	-.05	-.11
Number of unique documents	.20	.36
Highest rating of document use	.07	.11
Introduction	.00	.00
Conclusion	.15	.32
Type of writer	-.70	-.74

3.10, $p = .057$, indicating that the students' introductions and conclusions were not predictive of the persuasiveness. Finally, after Step 4, R^2 increased to .89, $F(3, 35) = 41.64$, $p = .000$, indicating that writing ability again contributed significantly to the overall prediction among the 11th graders.

Summary. The results from the multiple regressions indicate that the degree to which participants' reasoning was elaborated was strongly related to scores on overall quality and persuasiveness of the DBQ response, accounting for an average of 48% of the variance. Second, the frequency and degree to which participants used documents and integrated them into the fabric of their essays made a statistically significant contribution to their overall quality, about 25.5% on average. Third, the presence or absence of introductions and conclusions in a students' paper was important only for the eighth graders and was relatively unimportant at about 6%. Last, and not surprising, the type of writer (good vs. poor) affected the rating, contributing an additional average 26%. These findings are generally consistent with Ferretti and colleagues' (2009) results using persuasive prompts with younger students and add new information about the degree to which students' use of evidence contribute to the quality of their arguments.

Argumentative Strategies

Table 4 provides information regarding each topic and overall strategy use. In addition, close examination of participants' papers revealed that good writers at both grade levels consistently used one or more strategies for a given topic. For instance, when writing on the Progressive Era topic, all of the better 11th-grade writers included mention of values—for example, "Washington called for the development of a skillful, intelligence, diligent black community"—and 82% of the better 8th-grade writers used consequence, as in "If America did not stand up to Mexico to defend her territory, there would have been many disagreements." Other strategies were used far less frequently. For example, argument from analogy was used only 15% of the time, even among good 11th-grade writers. The following strategies were used across both topics: rule, verbal classification, example, consequence, commitment, inconsistent commitment, expert opinion, and cause to effect. Because we found 22 different strategies across DBQ topics, we limit the remainder of our results to similarities and differences in strategy use among good and poor writers.

Mexican-American War. We used chi-square analysis to evaluate whether the type of strategies participants used were related to the writer's ability. Each of the following strategies was found to be significantly related:

Verbal classification: Pearson $\chi^2(1, N = 44) = 7.42, p = .006$, Cramer's $V = .41$; good writers did so 79% of the time, whereas poor writers did so 38% of the time.

Example: Pearson $\chi^2(1, N = 44) = 11.12, p = .001$, Cramer's $V = .50$; good writers did so 57% of the time, whereas poor writers did so 6% of the time.

Consequence: Pearson $\chi^2(1, N = 44) = 5.05, p = .025$, Cramer's $V = .34$; good writers did so 82% of the time, whereas poor writers did so 50% of the time.

Expert opinion: Pearson $\chi^2(1, N = 44) = 5.21, p = .023$, Cramer's $V = .34$; good writers did so 61% of the time, whereas poor writers did so 25% of the time.

Fear: Pearson $\chi^2(1, N = 44) = 17.34, p = .000$, Cramer's $V = .63$; good writers did so 71% of the time, whereas poor writers did so 6% of the time.

Nonfunctional statements: Pearson $\chi^2(1, N = 44) = 12.16, p = .000$, Cramer's $V = .53$; good writers did so none of the time, whereas poor writers did so 38% of the time.

In contrast, 11 argumentation strategies were not used at reliably different rates by students who differed in writing ability—in part because several were used rarely, by one or two individuals. Participants did not use 6 strategies (sign, authority, distress, precedent, position to know, and agent from a past action) when writing about the Mexican-American War. These results show that in every case except nonfunctional statements, good writers used far more of these strategies than poor writers. The poor writers used more nonfunctional statements than the good writers.

Progressive Era. A second chi-square analysis was conducted to evaluate whether the type of strategies participants used were related to the writer's ability. Each of the following strategies was found to be significantly related:

Commitment: Pearson $\chi^2(1, N = 47) = 6.80, p = .009$, Cramer's $V = .38$; good writers did so 70% of the time, whereas poor writers did so 29% of the time.

Expert opinion: Pearson $\chi^2(1, N = 47) = 15.35, p = .000$, Cramer's $V = .57$; good writers did so 97% of the time, whereas poor writers did so 50% of the time.

Values: Pearson $\chi^2(1, N = 47) = 10.31, p = .001$, Cramer's $V = .47$; good writers did so all of the time, whereas poor writers did so 71% of the time.

In contrast, 13 strategies were not used at reliably different rates by students who were identified as good or poor writers, although additional strategies approached significance (e.g., 73% of the better writers included examples, in comparison to 43% of the weaker writers). In addition, 11th-grade writers in general used the consequence strategy quite often, as well as verbal classification and cause to effect, and they used rule, reference to authority, agent from a past action with less frequency and almost never included nonfunctional text. Participants did not use 7 strategies (slippery slope, fear, danger, distress, perception, moral justification, and analogy) when writing on the Progressive Era topic.

Summary. Our analyses of participants' use of argumentation strategies revealed both consistency and variability across writing ability and topics. Good writers routinely used more argumentation strategies than poor writers, and they used several strategies more consistently. Moreover, good writers used three strategies in particular (argument from example, argument from consequence, and argument from expert opinion) not only to warrant their standpoints about both topics but also to frame their use of evidence:

Washington's main idea for conditions to become better would be for African Americans to earn their way into society. They would have "to make himself, through is skill, intelligence, and character . . . [a] value to the community." This at least states that blacks do have character, skill, and intelligence, but it also says how they still have to push their way into society. Washington also said that in order to earn respect African Americans should learn "to produce what other people wanted and must have." I think this is totally opposite in what ex-slaves wanted. They didn't want to have to work for their respect, they wanted to already have it and to fit into society as easily as possible. Since white people weren't going to give them respect as quickly and easily, why should they try hard to earn it?

Yet the material in the DBQs also influenced the types of argument strategies that participants invoked. When writing about the Mexican-American War, both good and poor writers frequently used the cause-to-effect strategy (e.g., "I think that the U.S. government should have gone to Mexico. I think so because if the U.S. hadn't than California might have not been part of the U.S. government"). When writing on the Progressive Era topic, good and poor writers frequently used verbal classification (e.g., "Dubois infuriated the white race, rather than persuading them, through his demands of voting rights, elimination of discrimination, rights as freemen, enforced laws, and higher education"), consequence, and cause to effect (see Table 4 for additional examples). We did not find a "typical" approach to the constellation of

argumentative strategies, even among students who decided to respond to the historical question in similar ways (e.g., that DuBois was the better leader for African Americans at the time).

Length. A 2 (ability) \times 2 (grade) analysis of variance was used to evaluate the relationship between type of writers and length of essay. Table 5 presents descriptive information. Statistical analyses showed main effects for grade, $F(1, 87) = 97.25, p = .000$, and ability level, $F(1, 87) = 191.05, p = .000$. The interaction between ability group and grade was also significant, $F(1, 87) = 21.14, MSE = 4,795.53, p = .000$. Better writers at the 11th grade wrote the longest papers, and poor writers at the 8th grade wrote the shortest papers. In addition, good 8th-grade writers wrote longer papers than poor 11th-grade writers.

Discussion

We addressed two broad questions in this study: (a) How do 8th- and 11th-grade novices structure their essays when writing in response to DBQs about historical controversies? Do novices use everyday argumentative strategies when writing their arguments? If so, what argumentative strategies do they use? (b) What patterns in students' written responses are common between 8th- and 11th-grade students as they attempt to use historical evidence from documents in their writing? What patterns are common between students who are strong writers in comparison to those who are weak writers as they attempt to use historical evidence from documents in their writing? To accomplish these aims, we compared students' written historical arguments from two grade levels, which were composed in response to a standards-based, grade-appropriate DBQ involving a historical controversy from an available data set from a larger study (De La Paz et al., 2011). In doing so, we attempted to discern differences in the students' use of evidence, the structure of their written arguments, and the kinds of argumentation strategies within grade and due to writing ability. Because these analyses were conducted with an existing set of data, we selected writers for the current analysis using information about the overall persuasive quality of their papers from the larger study, and we took steps to limit our analyses to students who demonstrated adequate comprehension of the documents.

Findings Associated With Students' Grade and Ability Level

In the current study, we analyzed historical arguments from students who were asked to read primary source excerpts and write an essay. The topics were based on the curriculum but administered prior to instruction. While it should not be surprising that older and better writers generally wrote

longer arguments, cited documents more often, and integrated more unique documents to warrant their standpoints about historical controversies, this exploration of students' responses to DBQ tasks provides new information about students' abilities at more than one grade level. As such, we report how older and better writers provided more sophisticated use of evidence than younger and less able writers. Good writers chose more (and more relevant) evidence and went beyond citing their source of the evidence by explaining how their evidence was linked to their claim. Older writers demonstrated the greatest proficiency by balancing selected quotes, by evaluating the evidence, or by situating their evidence in the context of other content to advance their arguments. Finally, on two of three measures, the use of evidence in arguments written by older but weaker writers appeared similar to those written by younger better writers.

Furthermore, analyses showed that older and better students wrote essays with more highly elaborated structures. In fact, on average, the way and degree to which students elaborated their arguments contributed almost 50% of their reported level of quality. These findings replicate those reported by Ferretti et al. (2009), who used this approach to analyze the effects of goal structures on the argumentative writing of students who wrote about a topic related to school. Our results also extend Ferretti and colleagues' findings to disciplinary writing by showing that the structure of their arguments and the use of evidence were predictive of the overall quality of their historical arguments. Younger students who were asked a generic persuasive question (and who relied only on brainstormed ideas) considered up to 7 argumentative strategies on one topic. In contrast, in the current study, students considered 22 strategies across the two topics (14 were common to both). Finally, students' use of evidence accounted for a substantively significant proportion of the variance: an average of 25% of the overall quality of their written arguments, suggesting that the disciplinary nature was evident in their writing.

Importance of Topic

The analyses of argumentation strategies showed that their use was related to students' writing ability and the historical topics about which they wrote. On one hand, good writers used many more argumentation strategies than poor writers, and they frequently used some of the same strategies for more than one topic. We previously mentioned that *argument from example* and *argument from expert opinion* were frequently used by good writers to warrant their standpoints. In contrast, some argumentative strategies were invoked for specific topics (e.g., *argument from values* emerged only for the Progressive Era topic). So, while good writers clearly used more argumentation strategies

than poor ones, the historical topic clearly influenced the students' use of argumentation strategies. What accounts for the effect of topic?

Since Aristotle (1939), it has been theorized that interlocutors use specific argumentation strategies to accomplish particular rhetorical purposes (van Eemeren & Grootendorst, 2004; Walton, 1996; Walton et al., 2008). In many academic situations and certainly in the case of teacher-guided historical inquiry, students' discursive purposes are shaped by the teacher's questions and the documentary evidence made available to answer these questions. In the current study, students were asked to write argumentative essays in response to questions about two historical controversies. We illustrate the influence of the historical question and documentary evidence available to them by considering the Mexican-American War topic: "*Did the U.S. government have a reasonable (or unreasonable) argument for going to war with Mexico?*"

In this controversy, students had to argue about the reasonableness of the U.S. government's argument for going to war with Mexico, after reading four documents (and using one map to situate the conflict) that recounted the arguments of political leaders who either supported or opposed the war. These primary sources described the events in the topic and gave different authors' perspectives about the factors that precipitated the war, including the violation of American and Mexican borders, illustrations of these violations, the loss of American and Mexican life and land as a result of the purported Mexican hostilities, and the consequences for Mexicans and Americans of the purported encroachments. Our analyses showed that good writers tended to use argument from fear and consequences because many of the documents described appeals to fear and the purported adverse consequences that resulted from American and Mexican transgressions preceding the war. Furthermore, good writers often used argument from expert opinion because the source documents described the perspectives of four political leaders who presumably possessed special expertise about the conditions that precipitated the war. Finally, much of the political rhetoric used in these documents was vague and intended to elicit the audience's fears and passions. These features are well illustrated by President Polk's argument in favor of war: "After reiterated menaces, Mexico has passed the boundary of the United States, has invaded our territory and shed blood upon the American soil."

Impact of Writing Ability on Use of Argument Strategies

Our analyses also afforded us an opportunity to recognize similarities between good and poor writers, leading to potential implications for instruction. First, if one considers whether the students' argumentative strategies were adequate or reasonable with respect to the aim of persuasion, the better writers in our

sample used strategies that were based on facts and evidence from the documents more often than weaker writers. In contrast, weaker writers often relied on less precise strategies, such as *verbal classification*, *sign*, and *consequence*, and included nonfunctional content in their papers. Yet within each topic, all writers incorporated some strategies, which may be interpreted as a sign of emerging competence or a basis to focus initial efforts at instruction when designing an intervention for struggling writers. To illustrate, the *cause-to-effect strategy* appears to be understood in a nascent form by good and poor writers at both grade levels. If that is indeed true, it may be helpful to explicitly model and label this strategy for students and then (a) use this example as a bridge for teaching other argumentative strategies to students more formally in discussions (cf., Felton & Kuhn, 2001; Nussbaum & Edwards, 2011) that precede planning or composing activities (Wissinger & De La Paz, 2012) or (b) use this form of reasoning as students revise their essays.

Disciplinary Thinking

The hierarchical linear modeling results in our data provide compelling evidence regarding the contribution of disciplinary thinking to the overall quality rating when students are asked to compose arguments about historical controversies. In our sample, we noticed a range of ability between 8th and 11th graders, as younger, poor writers did not routinely cite documents in their essays. Yet, better 8th-grade writers regularly engaged in sourcing, and their evidence was rated as accurate or contextually relevant. Older, better writers' routinely explained or interpreted quotes and used quotes to substantiate their claims. Moreover, on the Progressive Era topic, the evidence that they selected was rated as balanced or significant. Interestingly, weaker (but older) writers often engaged in sourcing, but their evidence might have been misunderstood or inaccurate. Our findings corroborate and extend those provided by Young and Leinhardt (1998), who also used primary historical documents in an argumentative writing task, in that many good writers at both grade levels in the current study used more than one document to make a claim. We also noted that better 11th-grade writers used multiple documents (more than three, on average) to write their interpretation, something that was not reported in this landmark study with talented 10th graders. Finally, we noted several examples of contextualization and corroboration of evidence among students in both grade levels that Monte-Sano (2010) found in her sample of 11th-grade students.

To illustrate, consider the following essay by an eighth-grade good writer. Although this student did not provide an exceptionally strong example regarding contextualization in her paper (eighth-grade students tended to do well with

one aspect of their papers rather than both) and her language is not always fully developed, this student clearly understands more than one author's perspective and is able to corroborate their ideas into a clear argument:

The United States and Mexico went to war in 1846. I believe the U.S. Government did not have a reasonable argument for going to war with Mexico for many reasons. First of all, I agree that it was necessary to defend the boundary of the nations, but is it needed to go to war for that? Secondly, Abraham Lincoln believed that President Polk is wrong and shouldn't have gone to war. I agree with him. President Polk needed to find evidence. Thirdly, I don't believe Swain when he says to "redeem the Mexican people from...tyranny ...and to facilitate the entire removal of those rivals from this continent." By doing that, the United States is stealing territory and "violated the sovereignty of nations." It also seems like the Mexican people don't need saving from tyranny. I also don't think the Mexicans will be ruled by monarchy from Europe.

In conclusion, the U.S. government didn't have a good reason for going to war with Mexico. It was not reasonable.

Limitations

We acknowledge that certain limitations inherent in descriptive studies apply to the present study. Causal inference is not possible in our conclusions, due to our sampling method and lack of a clear intervention. Second, we realize some readers might contend that performance differences between 8th and 11th graders are confounded with topic differences caused by the standards-based focus of the DBQs administered at different grade levels. We disagree, based primarily on the remarkably similar and consistent hierarchical linear regression results across topics, and we point to the ecological validity in selecting topics that were relevant to students' curriculum; however, we cannot conclusively state that the differences that we attribute to grade are not potentially attributable to topic or sources. In addition, while we find the emerging strengths of the younger good writers to be especially heartening (who presumably knew less about writing historical arguments than the high school students), we nevertheless saw room for improving the quality of many students' essays. Our results also suggest that younger poor writers would benefit from learning more generally how to organize their ideas and the importance of including introductions and conclusions in their essays.

Conclusions

Writing argumentative essays in response to historical controversies is a complex endeavor that does not in itself promote disciplinary thinking (Grant, Gradwell, & Cimbricz, 2004). Recent work (cf. Monte-Sano, 2010) provide a compelling description of how high school students use evidence in their writing, and the findings from the current study augment those results by providing a detailed account of the ways that students use evidence and organize their arguments and the extent to which they use both general and disciplinary strategies in response to historical questions. Our results provide compelling evidence that students' use of general and disciplinary argumentation strategies relate not only to the historical topic and primary sources but also to differences in students' background characteristics (writing ability and grade level). We are encouraged by the results of our analyses that indicate that writers at both earlier and later secondary grades (and to some extent, good and poor writers within each grade) show similar patterns that may be seen as entry levels of performance to build on during history instruction. We also find grade-related patterns of document use that replicate earlier findings (e.g., Leinhardt, 2000; Young & Leinhardt, 1998) that show how older and better writers use multiple documents to create an overall interpretation and to corroborate and contextualize evidence in their essays. Future intervention research may explore argumentative strategies that are relevant to common historical topics so that teachers can be explicit about these strategies and students can become more facile at constructing arguments about the controversies they are attempting to reconcile.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

This study was made possible by funding for a Teaching American History grant (award no. U215X030180) from the U.S. Department of Education.

Note

1. When selecting students, we initially chose those who met target criteria regarding their reading and writing abilities and those who wrote at both the beginning and end of the school year. Had we realized the need to restrict our sample to the beginning of the year sooner, we would have been able to include other students who completed work at the beginning of the year but not the end.

References

- Anderson, R., Chinn, C., Chang, J., Waggoner, M., & Yi, H. (1997). On the logical integrity of children's arguments. *Cognition and Instruction, 15*, 135-167.
- Aristotle. (1939). *Topics*. Cambridge, MA: Harvard University Press.
- Bain, R. B. (2005). "They thought the world was flat?" Applying the principles of how people learn in teaching high school history. In National Research Council (Ed.), *How students learn: History, mathematics, and science in the classroom* (pp. 179-214). Washington, DC: National Academies Press.
- Bain, R. (2006). Rounding up unusual suspects: Facing the authority hidden in the history classroom. *Teachers College Record, 108*(10), 2080-2114.
- Britt, M. A., & Aglinskis, C. (2002). Improving students' ability to identify and use source information. *Cognition and Instruction, 20*, 485-522.
- Brophy, J., & Alleman, J. (2002). Primary-students' knowledge and thinking about the economics of meeting families' shelter needs. *American Educational Research Journal, 39*, 423-468.
- Brophy, J., & Alleman, J. (2003). Primary-students' knowledge and thinking about the supply of utilities (water, heat, and light) to modern homes. *Cognition and Instruction, 21*, 79-112.
- Christie, F., & Dreyfus, S. (2007). Letting the secret out: Successful writing in secondary English. *Australian Journal of Language & Literacy, 30*, 235-247.
- Coffin, C. (2006). *Historical discourse: The language of time, cause, and evaluation*. London: Continuum.
- Collingwood, R. G. (1943). *The idea of history*. Oxford, UK: Clarendon.
- Common Core State Standards Initiative. (2010). *Common core state standards for English language arts and literacy in history/social studies, science, and technical subjects*. Retrieved from http://www.corestandards.org/assets/CCSSI_ELA%20Standards.pdf
- 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*, 137-156.
- De La Paz, S., & Felton, M. (2010). Reading and writing from multiple sources in history: Effects of strategy instruction. *Contemporary Educational Psychology, 35*, 174-192.
- De La Paz, S., Malkus, N. N., Monte-Sano, C., & Montanaro, B. (2011). Evaluating American history teachers' professional development: Effects on student learning. *Theory and Research in Social Education, 39*, 494-540.
- Donovan, M. S., & Bransford, J. D. (2005). *How students learn: History, mathematics, and science in the classroom*. Washington, DC: National Academy Press.

- Educational Testing Service. (2004, November). *California Standardized Testing Program technical report, spring 2004 administration*. Sacramento: California Department of Education.
- Eisenberg, A., & Garvey, C. (1981). Children's use of verbal strategies in resolving conflicts. *Discourse Processes, 4*, 149-170.
- Felton, M., & Herko, S. (2004). From dialogue to two-sided argument: Scaffolding adolescents' persuasive writing. *Journal of Adolescent and Adult Literacy, 47*(8), 672-683.
- Felton, M., & Kuhn, D. (2001). The development of argumentative discourse skill. *Discourse Processes, 32*(2/3), 135-153.
- Ferretti, R. P., De La, & Paz, S. (2011). On the comprehension and production of written texts: Instructional activities that support content-area literacy. In R. O'Connor & P. Vadasy (Eds.), *Handbook of reading interventions* (pp. 326-355). New York: Guilford.
- Ferretti, R. P., Lewis, W. E., & Andrews-Weckerly, S. (2009). Do goals affect the structure of students' argumentative writing strategies? *Journal of Educational Psychology, 101*(3), 577-589.
- Ferretti, R. P., MacArthur, C. A., & Dowdy, N. S. (2000). The effects of an elaborated goal on the persuasive writing of students with learning disabilities and their normally achieving peers. *Journal of Educational Psychology, 92*, 694-702.
- Ferretti, R. P., MacArthur, C. A., & Okolo, C. M. (2001). Teaching for historical understanding in inclusive classrooms. *Learning Disability Quarterly, 24*, 59-71.
- Ferretti, R. P., MacArthur, C. A., & Okolo, C. M. (2005). Misconceptions about history: Reflections on teaching for historical understanding in an inclusive fifth-grade classroom. In T. E. Scruggs & M. A. Mastropieri (Eds.), *Advances in learning and behavioral disabilities* (Vol. 18, pp. 261-299). Oxford, UK: Elsevier Science / JAI Press.
- Ferretti, R. P., & Okolo, C. M. (1996). Authenticity in learning: Multimedia design projects in the social studies for students with disabilities. *Journal of Learning Disabilities, 29*, 450-460.
- Fránquiz, M. A., & Salinas, C. S. (2011). Newcomers developing English literacy through historical thinking and digitized primary sources. *Journal of Second Language Writing, 20*, 196-210.
- Graham, S., & Perin, D. (2007). A meta-analysis of writing instruction for adolescent students. *Journal of Educational Psychology, 99*, 445-476.
- Grant, S. G., Gradwell, J. M., & Cimbricz, S. K. (2004). A question of authenticity: The document-based question as an assessment of students' knowledge of history. *Journal of Curriculum and Supervision, 19*, 309-337.
- Halldén, O. (1997). Conceptual change and the learning of history. *International Journal of Educational Research, 27*, 201-210.
- Hexter, J. H. (1971). *The history primer*. New York: Basic Books.

- Holt, T. (1990). *Thinking historically: Narrative, imagination, and understanding*. New York: College Entrance Examination Board.
- Husbands, C. (1996). *What is history teaching? Language, ideas and meaning in learning about the past*. Buckingham, UK: Open University Press.
- Kuhn, D. (1991). *The skills of argument*. New York: Cambridge University Press.
- Leinhardt, G. (2000). Lessons on teaching and learning in history from Paul's pen. In P. Stearns, P. Seixas, & S. Wineburg (Eds.), *Knowing, teaching and learning history* (pp. 223-245). New York: New York University Press.
- Lewis, W. E., & Ferretti, R. P. (2011). Topoi and literary interpretation: The effects of a critical reading and writing intervention on high school students' analytic literary essays. *Contemporary Educational Psychology, 36*(4), 334-354.
- Mink, L. (1987). *Historical understanding*. Ithaca, NY: Cornell University Press.
- Moje, E. (2008). Foregrounding the disciplines in secondary literacy teaching and learning: A call for change. *Journal of Adolescent and Adult Literacy, 52*(2), 96-107.
- Monte-Sano, C. (2008). Qualities of historical writing instruction: A comparative case study of two teachers' practices. *American Educational Research Journal, 45*, 1045-1079.
- Monte-Sano, C. (2010). Disciplinary literacy in history: An exploration of the historical nature of adolescents' writing. *Journal of Learning Sciences, 19*(4), 539-568.
- Monte-Sano, C., De, La, & Paz, S. (2012). Using writing tasks to elicit adolescents' historical reasoning. *Journal of Literacy Research, 44*(3), 273-299.
- Newell, G. E., Beach, R., Smith, J., & VanDerHeide, J. (2011). Teaching and learning argumentative reading and writing: A review of research. *Reading Research Quarterly, 46*(3), 273-304.
- Newmann, F. (1990). Higher order thinking in teaching social studies: A rationale for the assessment of classroom thoughtfulness. *Journal of Curriculum Studies, 22*(1), 41-56.
- 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*, 492-504.
- Nussbaum, E. M., & Edwards, O. V. (2011). Critical questions and argument strata-gems: A framework for enhancing and analyzing students' reasoning practices. *Journal of the Learning Sciences, 20*, 443-488.
- Perfetti, C. A., Britt, M. A., & Georgi, M. C. (1995). *Text-base learning and reasoning: Studies in history*. Hillsdale, NJ: Erlbaum.
- Salahu-Din, D., Persky, H., & Miller, J. (2008). *The nation's report card: Writing 2007* (NCES Document No. 2008-468). Washington, DC: National Center for Education Statistics.

- Schleppegrell, M. J., Achugar, M., & Oteiza, T. (2004). The grammar of history: Enhancing content-based instruction through a functional focus on language. *TESOL Quarterly*, 38(1), 67-93.
- Shanahan, T., & Shanahan, C. (2008). Teaching disciplinary literacy to adolescents: Rethinking content-area literacy. *Harvard Educational Review*, 78(1), 40-59.
- Shemilt, D. (1983). The devil's locomotive. *History and Theory*, 22, 1-18.
- Stein, N., & Miller, C. (1993). The development of memory and reasoning skill in argumentative contexts: Evaluating, explaining, and generating evidence. In R. Glaser (Ed.), *Advances in instructional psychology* (pp. 285-335). Hillsdale, NJ: Erlbaum.
- Stevens, R., Wineburg, S., Herrenkohl, L. R., & Bell, P. (2005). Comparative understanding of school subjects: Past, present, and future. *Review of Educational Research*, 75, 125-157.
- Suppe, F. (1998). The structure of a scientific paper. *Philosophy of Science*, 65, 381-405.
- Van Drie, J., & Van Boxtel, C. (2008). Historical reasoning: Towards a framework for analyzing students' reasoning about the past. *Educational Psychology Review*, 20, 87-110.
- van Eemeren, F. H., & Grootendorst, R. (1992). *Argumentation, communication, and fallacies: A pragma-dialectical perspective*. Mahwah, NJ: Erlbaum.
- van Eemeren, F. H., & Grootendorst, R. (2004). *A systematic theory of argumentation: The pragma-dialectical approach*. Cambridge, England: Cambridge University Press.
- van Eemeren, F. H., Grootendorst, R., & Henkemans, F. S. (1996). *Fundamentals of argumentation theory: A handbook of historical backgrounds and contemporary developments*. Mahwah, NJ: Erlbaum.
- van Eemeren, F. H., Grootendorst, R., & Henkemans, F. S. (2002). *Argumentation: Analysis, evaluation, and presentation*. Mahwah, NJ: Erlbaum.
- VanSledright, B. (2002). Confronting history's interpretive paradox while teaching fifth graders to investigate the past. *American Educational Research Journal*, 39(4), 1089-1115.
- von Ranke, L. (2010). *The theory and practice of history*. Florence, KY: Routledge.
- Walton, D. N. (1996). *Argumentation schemes for presumptive reasoning*. Mahwah, NJ: Erlbaum.
- Walton, D., Reed, C., & Macagno, F. (2008). *Argumentation schemes*. Cambridge, England: Cambridge University Press.
- Wineburg, S. (1991a). Historical problem solving: A study of the cognitive processes used in the evaluation of documentary and pictorial evidence. *Journal of Educational Psychology*, 83(1), 73-87.
- Wineburg, S. (1991b). On the reading of historical texts: Notes on the breach between school and the academy. *American Educational Research Journal*, 28, 495-519.

- Wineburg, S., & Martin, D. (2009). Tampering with history: Adapting primary sources for struggling readers. *Social Education, 73*(5), 212-216.
- Wissinger, D., & De La Paz., S. (2012, April). *Effects of genre and content knowledge on historical thinking with academically diverse high school students*. Paper presented at the annual meeting of American Educational Research Association, Vancouver, BC, Canada.
- Young, K. M., & Leinhardt, G. (1998). Writing from primary documents: A way of knowing in history. *Written Communication, 15*, 25-68.
- Zwiers, J. (2006). Integrating academic language, thinking, and content: Learning scaffolds for non-native speakers in the middle grades. *Journal of English for Academic Purposes, 5*, 317-332.