LEARNING, INSTRUCTION, AND COGNITION

Effects of Genre and Content Knowledge on Historical Thinking With Academically Diverse High School Students

Susan De La Paz
University of Maryland

Daniel R. Wissinger
Indiana University of Pennsylvania

Historians use a range of genres in presenting their subjects, yet educators have increasingly privileged argumentation to help novices to reason with historical content. However, the influence genre and content knowledge are relatively unmeasured in this discipline. To learn more, the authors asked 101 eleventh-grade students to compose an argument or a summary on the Gulf of Tonkin and analyzed basic and disciplinary reading and writing measures. Results indicate that students with adequate content knowledge performed better on a disciplinary reading measure when composing arguments, and students with limited content knowledge demonstrated greater comprehension when composing summaries. Students with more knowledge wrote more; however, students who wrote summaries were not disadvantaged in terms of level of historical thinking or overall quality. Last, providing students with disabilities with a simple reading accommodation afforded them the ability to participate in the disciplinary literacy task at levels comparable to their peers.

Keywords disciplinary literacy, content learning, high school, genre, writing

TODAY’S WRITING RESEARCHERS consider cognitive processing models (Hayes & Flower, 1980), what it means to write in the content areas (Bangert-Drowns, Hurley, & Wilkinson, 2004), disciplinary goals and processes that professionals engage in (Shanahan & Shanahan, 2008) as well as discourse communities for whom people write (Langer, 1992; Stockton, 1995). When writing about the past, historians “interpret and integrate information from different sources . . . to define a problem, speculate about alternative actions, and reformulate information in supporting a point of view” (Greene, 1994, p. 90). Causality and explanation are central to writing in history.
(Coffin, 2004; Leinhardt, Stainton, & Virji, 1994) as are the means by which historians connect claims with documentary evidence (Collingwood, 1943; Monte-Sano, 2010; Wineburg, 1991). For example, stating where evidence comes from (i.e., sources of quotations and information) allows others to understand and evaluate the basis for the writer’s claim (Hexter, 1971).

Although historians use a variety of rhetorical devices and genres in presenting their subjects (Beaufort & Williams, 2005), dominant forms of historical discourse include narrative and argument (Coffin, 2004, 2006; Marius & Page, 2005). Barbara Tuchman, author of *The Guns of August* (1962), describes an “inherent validity” (p. 70) of the narrative genre, describing it as central to understanding causation. “When [events] are arranged in sequence as strictly as possible, down to the week and day, sometimes even time of day, cause and effect which may have been previously obscure will become clear” (Tuchman, 1981, p. 70). British historian and historiographer John Tosh speaks in more general terms, noting that historical writing takes on many forms, depending on whether an author’s purpose is “to re-create the past [or whether he wishes] to interpret it” (1984, p. 95). However, the most essential criteria in determining the quality of historical writing is the provision of evidence, and historians interpret evidence when crafting narratives and negotiate between competing narratives in determining what should be told (Levinsohn, 2010).

Educational reformers promote domain-specific reasoning tasks for adolescent learners, by prioritizing connections between literacy and content learning within subjects (Moje, 2008; Young & Leinhardt, 1998). Advocates for a disciplinary approach to literacy suggest that adolescents attempt reading and writing tasks in ways that are based on observations on how experts think, observing differences in reasoning and problem solving across the sciences (e.g., biology; Carter, Ferzli, & Wiebe, 2007), mathematics (Brown, 2007), and history (Monte-Sano, 2010; National Research Council, 2000). For example, in history, experts simultaneously judge the influence of time, place, and cultural and linguistic norms that existed in the time period in question on the motivations of the authors (Wineburg, 2001). Students will likely benefit from engaging in similar processes as they work to construct an evidence-based interpretation of past events (Britt & Aglinskas, 2002; Nokes, Dole, & Hacker, 2007).

**Reading and Writing in History Classrooms**

In developing novices’ historical thinking, the kinds of texts with which students work have been shown to influence how they attempt to reason in history. For example, students are more likely to think analytically and interact with texts if they read primary documents (Rouet, Britt, Mason, & Perfetti, 1996). However, research in this area has revealed cognitive limitations in students (e.g., ignoring rather than reconciling conflicting information; Stahl, Hynd, Britton, McNish, & Bosquet, 1996) and potential for students to produce deeper conceptual understandings (Britt & Aglinskas, 2002; Spiro, Coulson, Feltoovich, & Anderson, 1994; Wiley & Voss, 1999).

Writing essays from documents can facilitate students’ ability to learn historical content (Bangert-Drowns et al., 2004; De La Paz, 2005; Smith & Niemi, 2001). However, to date, there are inconsistent findings on how genre influences how novice thinkers develop conceptual understandings within the disciplines. Writing arguments has been shown to promote better audience awareness and syntactic complexity than writing narratives and description (Crowhurst & Piche, 1979) and help advanced students to construct deep understanding of content (Wiley & Voss, 1999). Stahl and colleagues (1996) found that students who wrote opinions used more
ideas that came from multiple texts when compared with students who wrote descriptions. Moreover, there seems to be a connection between writing arguments and greater attention to source information (Le Bigot & Rouet, 2007; Stahl et al., 1996; Wiley & Voss, 1999). In contrast, writing argumentative essays, in themselves, do not always promote disciplinary thinking (cf. Grant, Gradwell, & Cimbricz, 2004). Greene (1994) found that students sometimes respond to requests to create arguments by using documents to support specific standpoints, thus neglecting to fully considering the evidence within the sources and Monte-Sano and De La Paz (2012) found varying results depending on specific foci of an argumentative writing prompt (whether students were encouraged to engage in sourcing, corroboration, or causation as opposed to writing in first person).

Comparisons of narrative and argumentative writing among adolescents are rare, especially when asked to reconcile conflicting historical accounts. Wiley and Voss (1998) found that students who were assigned to write arguments gained deeper text comprehension and produced more transformed and integrated essays than did students assigned to write summaries, narratives, or explanations on the same topic. In contrast, narrative writing in schools is typically associated with writing fiction and the development of plot, character, and so forth (Common Core State Standards, 2010). This schism may exist in part because historians who write narratives are motivated by a different purpose (e.g., responding to a historical question), thus demonstrating command of a complex form of narrative (Beaufort & Williams, 2005).

Kiuhara, Graham, and Hawken (2009) noted that history teachers often assign adolescents to write summaries. This may be beneficial as a meta-analysis of writing instruction (Graham & Perin, 2007) revealed distinct advantages for summary writing, perhaps because this genre requires that students understand and organize information, then condense it into a shorter, more concise version of the original text. Doing so means students must first distinguish between what is important and what is not, requiring conscious thought, judgment, and metacognitive effort (Taylor, 1982). According to Rinehart, Stahl, and Erikson (1986), summarization can improve students’ long-term retention of information, leading to positive effects on their content learning. Moreover, teaching students to summarize can improve their comprehension (Duke & Pearson, 2002; Graham & Perin, 2007; Taylor & Beach, 1984). Some studies indicate that summary writing can be more facilitative than argumentative writing tasks when working with multiple documents (Gil, Bråten, Vidal-Abarca, & Strømsø, 2010a, 2010b). While critics warn that writing summaries may encourage students to passively transfer ideas from documents to their essays (Le Bigot & Rouet, 2007; Wiley & Voss, 1999), we lack information on the effects of summarization on students’ writing with multiple primary sources in history.

Disciplinary tasks place cognitive demands on students while reading historical documents, especially because novices must build background knowledge about each historical event to integrate information from sources with mental representation of each event (De La Paz, 2005; Van Drie & Van Boxtel, 2008). In contrast, historians are able to read multiple documents fluently, simultaneously considering the relations within and among documents and the events described in each, as well as the entire situation described or discussed in the documents (Perfetti, 1999). In addition, historians’ adjust their schemas for reading documents in accordance with task demands, differentiating how to read familiar versus unfamiliar texts (Leinhardt & Young, 1996). However, a consistent finding across grade levels is that the degrees to which students know about the topic influences their ability to reason and comprehend with multiple documents (Rouet et al., 1996; Stahl et al., 1996; Young & Leinhardt, 1998).
Evidence from a few studies suggests that writing summaries can be more beneficial than can writing arguments when students read from documents that they are unfamiliar with (Bråten, Strømsø, & Britt, 2009; Gil et al., 2010; Le Bigot & Rouet, 2007). In these investigations, students who were identified as having low background knowledge performed more poorly when writing arguments than did students who wrote summaries, producing fewer transformations with less integration of ideas in their compositions. This point is important to consider when considering the needs of academically diverse students, especially those who struggle most with learning, as it is now the norm that they too are asked to attempt discipline-based reading and writing activities in the classroom.

Considering the instructional needs of struggling adolescents and students with learning disabilities in history classrooms is especially important because we lack clear guidelines on how to best facilitate these students’ disciplinary writing. Teachers will benefit from knowing how to teach students with learning disabilities who have general deficits in verbal learning and memory (cf. Mastropieri & Scruggs, 2010), limitations in their ability to consider multiple perspectives (Bouck, Okolo, Englert, & Heutsche, 2008), or in their ability to differentiate relevant from irrelevant details (Bulgren, Deshler, & Lenz, 2007), because these difficulties have definite implications for reasoning with historical content. It is especially important to learn whether students with limited historical knowledge can benefit from writing arguments as opposed to writing summaries, when working with primary source documents.

METHOD

Given this background, we asked the following questions: Does writing genre influence students’ understanding of and ability to write from multiple primary source documents? Second, do differences in students’ level of content knowledge before instruction influence their understanding of and ability to write from primary documents? With these purposes in mind, we asked four, eleventh-grade social studies teachers to implement a jointly developed lesson on the Gulf of Tonkin Incident to their students. We randomly assigned students to write either an argumentative or summary essay, using the same historical documents. We gave each student a comprehension test about the documents, after they wrote their essay, and examined their papers for additional evidence that they understood the documents. Last, we examined the effects of genre and content knowledge on a subsample of students who were identified with learning disabilities in the classrooms.

Setting and Participants

We recruited 101 eleventh-grade students in U.S. History courses from a large rural school district in south-central Pennsylvania. The majority (N = 91) included general education students, with about 10% (n = 10) identified with specific learning and/or behavioral disabilities. Fifty-four percent of the students were girls, 46% were boys. Reported ethnicities were 87% Caucasian, 7% African American, 4% Hispanic, and 2% Asian. Of the students, 23% met federal criteria as economically disadvantaged.
Table 1 presents the means and standard deviations for the characteristics of students. A series of one-way analysis of variance tests indicated that there were no significant differences among students assigned to the two writing genres in terms of gender, $F(1, 99) = .474, p = .49$; ethnicity, $F(1, 99) = .779, p = .38$; number of economically disadvantaged students, $F(1, 99) = .103, p = .75$; or number of students identified as having a disability, $F(1, 99) = .317, p = .56$. We performed additional analysis of variance tests to evaluate the relation between the two writing genres (argument and summary) and students’ writing ability (using scores from the writing portion of the Pennsylvania State System of Assessments (PSSA), a metric of the quality of students’ writing, and their Wechsler Individual Achievement Test (WIAT-III) essay subtest scores), reading ability (using standardized reading scores from the PSSA), and students’ most recent semester grades in social studies. There were no significant differences between the two genre groups on PSSA writing scores, $F(1, 90) = .188, p = .67$; WIAT-III essay composition scores, $F(1, 99) = 1.41, p = .24$; PSSA reading scores, $F(1, 99) = 3.42, p = .06$; or final social studies grades, $F(1, 90) = .227, p = .64$. However, there were significant differences in initial reading and writing abilities among students with and without disabilities on the WIAT-III, $F(1, 99) = 5.54, p = .02$ and PSSA reading scores, $F(1, 90) = 4.29, p = .04$.

Design

We used an experimental design with four sections of an 11th-grade U.S. History class. Students, including those with disabilities, were randomly assigned to either (a) argument or (b) summary condition in each of the four classes.

Materials

The topic under investigation provided multiple perspectives about the Gulf of Tonkin Incident and the resultant Gulf of Tonkin Resolution that was passed by the U.S. Congress on August 8, 1964. Before the study began, the four teachers developed a shared lesson plan to overview the two attacks on U.S. military vessels in the Gulf of Tonkin on August 2nd, and on August 4, 1964.
We relied on prior investigations (De La Paz & Felton, 2010; Stahl et al., 1996; Wiley & Voss, 1999) to guide us in preparing our primary and secondary source documents and determine when to administer content and comprehension measures (see the Appendix for document set) and developed materials with summary information from the school’s textbook (McDougall Littel, 2003) and adapted excerpts from Lyndon Johnson’s Midnight Address to the American Public, a statement by Senator Wayne Morse, and a New York Times newspaper article written the day after the Gulf of Tonkin Resolution was passed by Congress.

General Procedures

The four U.S. History teachers presented a jointly developed lesson with historical background information related to the topic in one 90-min class to ensure that students would all know similar information about the historical context for the subsequent historical reading and writing task. In this lesson, teachers provided an overview of the topic using a timeline and identified the major historical agents who were involved in the Gulf of Tonkin Incident and Resolution. The lesson ended with clarification from the teachers as to what was controversial about the historical events. The next day, students then completed the written expression subtest of the WIAT-III and a multiple-choice content measure about the Gulf of Tonkin Incident and Resolution. Next, students read the adapted primary documents about the Gulf of Tonkin Incident and Resolution and composed a historical argument or historical summary. Last, students completed a multiple-choice, literal reading comprehension test. A reading specialist at the school read the document set to the students with disabilities aloud before they read the primary sources and composed essays. This accommodation was outlined in the students’ individualized education plans and was regularly provided during state and local standards-based assessments (e.g., reading, writing, and math portions of the PSSA).

Writing Prompt

Before starting on the documents, students were presented with a page with the following writing task:

Historians work from sources including newspaper articles; autobiographies and documents like journal entries and public speeches to create histories. Your task is to take the role of historian and develop (an argument or a summary) about the question: Would you have voted for or against using force in response to the Gulf of Tonkin Incident in 1964? Read the documents and develop (an argument or a summary) that considers both sides of the issue in response to the Gulf of Tonkin Incident in 1964: If you were a member of Congress at the time this event unfolded, would you have voted for or against using force in response to the Gulf of Tonkin Incident?

Students were also instructed as follows: “Read through all the information before writing and develop an essay that considers both sides of the issue. Use evidence from the documents to support your writing.” They were allowed to read and reread the documents in any order. These directions were similar to the procedures outlined in the study by Gil and colleagues (2010), in which students were asked to read and work with multiple documents about a scientific topic (i.e., climate change) and asked to write an argument or summary. In contrast, it is important to note that students did not receive instruction in sourcing, corroboration or contextualization or
were supported as they attempted to reason with the documents. Last, students were advised to
take 20 min to read through the documents and 40 min to write either the argument or summary
essay. However, in contrast with previous work (Gil et al., 2010; Stahl et al., 1996), students were
able to look back through the documents while constructing written responses. The time limit of
20 min for reading and 40 min for writing was based on criteria outlined on the writing portion
of the PSSA for 11th-grade students.

Dependent Measures

Six dependent measures were used in this study: (a) one measure of content knowledge, of the
historical topic, (b) two measures of reading comprehension, and (c) three measures of writing
ability. A brief description of each follows.

Content Knowledge

Background information for the content knowledge assessment provided to students before
the start of the investigation came from the district-selected textbook, *The Americans*
(McDougal Littell, 2003). The four participating teachers created a 10-item, multiple-choice test, from their
instructional materials and the textbook to gauge students’ familiarity with the topic they were
exploring. Scores on the 10-item assessment were used to identify students as having either low
or high content knowledge. Grouping was based on whether scores were two standard deviations
($SD = 1.8$) above or below the mean ($M = 5.9$). Students who scored two standard deviations
below the average score of the sample were considered to have low levels of content knowledge
about the topic (4.0 and below). Similarly, students who scored two standard deviations above the
average score of the sample were considered to have high levels of content knowledge (8.0 and
above). Because there were some students who scored between these benchmarks, we decided
that individuals who correctly answered at least two of the three most difficult questions could also
be considered as having high levels of content knowledge. Using these criteria, 41 students appeared
to have low levels of content knowledge about the Gulf of Tonkin Incident and Resolution, while
60 students appeared to have high levels of content knowledge before completing the experimental
writing task.

Literal Comprehension

Students were asked to demonstrate what they understood about the Gulf of Tonkin Incident
and the Gulf of Tonkin Resolution by answering 12 multiple-choice questions after writing
their essays (Gil et al., 2010; Le Bigot & Rouet, 2007; Wiley & Voss, 1999). The second
author corrected the comprehension test and an independent rater scored a random subset of the
assessments (25%). Interrater reliability was 100%.

Transformations

A second, disciplinary measure of reading comprehension was based on the number and type
of idea units that students used from the documents in their essays (Wiley & Voss, 1999). Although
scoring this measure is based on examining students’ writing, Scardamalia and Bereiter (1987)
found that the type of idea units students transformed in their writing was a strong indicator of disciplinary text comprehension.

Using procedures outlined by Wiley and Voss (1996), students’ essays were segmented into idea units containing one or more related items of information. Several variations of this coding system have been used in more recent work and were followed in our coding scheme (Gil et al., 2010; Le Bigot & Rouet, 2007). Idea units in each of the student’s essays were divided into four categories: paraphrases, elaborations, additions, or misconceptions. In students’ essays, idea units were coded as \textit{paraphrases} if students used their own words without changing the meanings expressed in the documents (e.g., “The dispute over the proposed reaction to the Gulf of Tonkin Incident can best be summarized with two prime arguments; the pro military response side which states that the attack was unprovoked” [see Document 1] “and that the acts being committed by North Vietnam were a threat to the whole world” [see Document 3] “not just Southeast Asia”).

Idea units were coded as \textit{elaborations} if they included some information from the documents in combination with an idea that the student knew about the historical topic (e.g., “By giving Johnson the capability to use military action in North Vietnam the spread of communism” [see Document 1], “would pause and possibly diminish, so by granting Johnson with greater executive power through the Gulf of Tonkin Resolution, communism may eventually be gotten rid of”). Idea units were also coded as elaborations if the student combined two or more pieces of information, either within or across texts, which were not connected in the source (e.g., “Though I feel as though Congress may have had their intentions and hearts in the wrong place by voting for the Gulf of Tonkin Resolution less than three days after the attacks in the Gulf of Tonkin” [see Document 3], “they must have been concerned for the people of Southeast Asia” [see Document 2] “and their freedom”).

Idea units were coded as \textit{additions} if they contained only related information from prior knowledge or personal opinions about the Gulf of Tonkin Incident, the Gulf of Tonkin Resolution, or the Vietnam War (e.g., “However, in opposition to the argument of not responding to the attacks in the Gulf of Tonkin with force, I personally believe we would be better of risking a few to save many, than to risk nothing and save a few”). In addition, we supplemented the coding system used by Wiley and Voss (1996) with a category entitled \textit{misconceptions} (Gil et al., 2010). Idea units were coded as misconceptions if they contain false statements in relation to the content of the source material (e.g., “Another reason that the United States should have used force in response to the Incident in the Tonkin Gulf was that when the US military was attacked, some of our loved ones got hurt or even killed”). This statement was coded a misconception because, although U.S. Naval vessel were attacked in the Tonkin Gulf, the background information indicated that no casualties or injuries were reported. At the same time, idea units that contained minor mistakes (date, locations, etc.) without demonstrating lack of comprehension were coded as paraphrases, elaborations, or additions.

The second author counted the number of transformations and computed a total score by adding the number of paraphrases, elaborations, and additions and subtracting the number of misconceptions from the total sum score. Students were awarded two points for each elaboration and one point for each paraphrase or addition. Elaborations were scored a two because they reflected a greater degree of transformation than both paraphrases and additions (Gil et al., 2010; Wiley & Voss, 1996). Misconceptions were scored negatively as they reflected a lack of understanding of the document set. Next, a social studies teacher who was blind to the experimental
conditions in the study rated a random subset of summaries and argument essays (30%) using the same coding systems. The overall agreement was 86% with further breakdown as follows: agreement was 89% for paraphrases, 78% for elaborations, 97% for additions, and 83% for misconceptions.

Written Essays

Three measures were used to explore students’ written essays. These included (a) a disciplinary measure, of historical thinking; (b) a basic measure, of writing quality (using criteria from the PSSA); and (c) the length of students’ essays.

**Historical Thinking**

An analytic rubric was developed to capture the extent to which the students exhibited historical thinking in their essays, on the basis of prior work (Monte-Sano & De La Paz, 2012). Students’ essays were scored for presence of (a) substantiation, (b) perspective recognition, and (c) contextualization based on a 4-point scale (ranging from 0 to 3 on each element). Substantiation emphasized the extent to which students provided evidence and explanation in support of a standpoint. Perspective recognition focused on students’ skills in presenting the documents as authors’ point of view that. Contextualization outlined the degree to which students identified and integrated their arguments or summaries in the appropriate time, place, and setting—linking related events. These are all aspects of historical thinking that Monte-Sano (2010) found in adolescents’ history essays; as such it seemed to be a reasonable measure by which to compare the influence of genre and content knowledge for students in the present study. Two social studies secondary teachers received benchmark criteria for scoring student essays, sample student responses, and 10 essays to be used for scoring practice, and met for roughly 2–3 hr to practice scoring. Interrater reliability using a criterion of exact agreement was established at \( r = .80 \).

The raters each scored each essay independently and achieved the following level of reliability for each analytic trait (substantiation, \( r = .88 \); perspective, \( r = .89 \); and contextualization, \( r = .90 \)).

**Quality**

Each essay was scored separately using the PSSA scoring index. The PSSA quality index emphasizes five dimensions of effective writing: focus, content, organization, style, and conventions. The quality of the five dimensions was also based on a 4-point scale (PSSA, 2010). Focus highlighted making a single controlling point with an awareness of the genre (i.e., writing an argument, summary, narrative). Content emphasized the inclusion of ideas developed through facts, examples, anecdotes, details, opinions, reasons, and/or explanations. Organization referred to the development of the essay within and across paragraphs using transition devices and included introduction and conclusion paragraphs. Style outlined word choice, arrangement, and sentence structure that communicated the writers’ tone and voice. Conventions detailed the grammar, mechanics, spelling, usage, and sentence formation. Two high school teachers received benchmark criteria for scoring student essays, sample student responses, and 10 essays to be used for scoring practice, and met for roughly 2–3 hr to practice scoring. Interrater reliability...
using a criterion of exact agreement or adjacent scores was established at 80%. Final reliability on the complete set of essays was .913% on the PSSA total score, with further breakdown as follows: focus, $r = .92$; content, $r = .94$; organization, $r = .95$; style, $r = .91$; and conventions, $r = .98$.

**Length**

All essays were scored on the total number of words written. This number included all words that represented a spoken word regardless of spelling. A word processing program determined length.

**RESULTS**

Table 2 shows the means scores and standard deviations for literal reading comprehension test, number of transformations, historical thinking, writing quality, and essay length. Results for subgroups of students (those with low or high content knowledge and students with and without disabilities) are shown in Tables 3 and 4.

**Content Knowledge**

Initial differences among students in the argument and summary genre were not significant for content knowledge $F(1, 99) = .232, MSE = .776, p = .63$. Therefore, because there were no differences in content knowledge between the two writing genres, we decided to examine the influence of students’ content knowledge on all subsequent reading and writing outcome measures.

**Reading Comprehension Measures**

We conducted a one-way analysis of variance to evaluate the relation between the two genres and our literal reading comprehension test. There were no statistically significant differences $F(1, 101) = .219, MSE = .446, p = .64$. Thus, assigning students to write an argument or summary

<table>
<thead>
<tr>
<th>Writing condition</th>
<th>Argument</th>
<th>Summary</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M</td>
<td>SD</td>
</tr>
<tr>
<td>Comprehension</td>
<td>10.12</td>
<td>1.85</td>
</tr>
<tr>
<td>Transformation</td>
<td>11.90</td>
<td>5.48</td>
</tr>
<tr>
<td>Historical thinking</td>
<td>5.19</td>
<td>2.68</td>
</tr>
<tr>
<td>Writing quality</td>
<td>16.12</td>
<td>3.15</td>
</tr>
<tr>
<td>Number of words</td>
<td>171.62</td>
<td>73.93</td>
</tr>
</tbody>
</table>
about a historical topic (e.g., the Gulf of Tonkin Incident and the Gulf of Tonkin Resolution) did not have a significant effect on a measure of literal reading comprehension. In addition, we used a $2 \times 2$ analysis of variance to evaluate the relation between students with low levels of content knowledge, high levels of content knowledge, and assignment to argument or summary genre on multiple-choice reading comprehension scores. Table 3 presents descriptive information. The statistical analyses showed main effects for level of content knowledge, $F(1, 101) = 33.666, MSE = 68.032, p = .00$. Students with high levels of content knowledge performed significantly better than students with low levels of content knowledge on the literal reading comprehension test ($M = 10.8$ and $M = 9.15$, respectively). No main effects were found for writing genre, $F(1,101) = .132, MSE = .267, p = .72$. It is important to note that the interaction between level of content knowledge and writing genre was significant $F(1, 101) = 6.206, MSE = 12.542, p = .01$. This suggested that writing genre and level of content knowledge about a historical topic influenced students’ literal reading comprehension. Students with low levels of content knowledge about the historical topic who were assigned to write summaries performed better on the multiple-choice reading comprehension test. In contrast, students with high levels of content knowledge who were assigned to write arguments performed better on the multiple-choice reading comprehension test than their peers who wrote summaries.

<table>
<thead>
<tr>
<th>Measure</th>
<th>Argument M</th>
<th>Argument SD</th>
<th>Summary M</th>
<th>Summary SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Multiple choice comprehension measure</td>
<td>8.73</td>
<td>1.83</td>
<td>11.13</td>
<td>1.04</td>
</tr>
<tr>
<td>Transformation total score</td>
<td>10.68</td>
<td>5.02</td>
<td>12.80</td>
<td>5.71</td>
</tr>
<tr>
<td>Overall writing quality</td>
<td>15.32</td>
<td>2.75</td>
<td>16.70</td>
<td>3.33</td>
</tr>
<tr>
<td>Historical thinking rubric</td>
<td>5.36</td>
<td>2.34</td>
<td>5.07</td>
<td>2.94</td>
</tr>
<tr>
<td>Essay length</td>
<td>154.36</td>
<td>54.73</td>
<td>182.34</td>
<td>73.07</td>
</tr>
</tbody>
</table>

In TABLE 4, we present means and standard deviations for students with a learning disability and general education students for each writing task. The statistical analyses showed main effects for writing genre, $F(1, 101) = 132.75, MSE = 0.267, p = .00$. Students who wrote summaries performed significantly better than those who wrote arguments. No main effects were found for learning disability, $F(1,101) = .132, MSE = .267, p = .72$. It is important to note that the interaction between learning disability and writing genre was significant $F(1, 101) = 6.206, MSE = 12.542, p = .01$. This suggested that writing genre and learning disability about a historical topic influenced students’ literal reading comprehension. Students with learning disabilities who were assigned to write summaries performed better on the multiple-choice reading comprehension test. In contrast, students with learning disabilities who were assigned to write arguments performed better on the multiple-choice reading comprehension test than their peers who wrote summaries.

<table>
<thead>
<tr>
<th>Measure</th>
<th>Learning disability M</th>
<th>Learning disability SD</th>
<th>General education M</th>
<th>General education SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Multiple choice comprehension measure</td>
<td>9.17</td>
<td>2.23</td>
<td>10.24</td>
<td>1.79</td>
</tr>
<tr>
<td>Transformation total score</td>
<td>13.00</td>
<td>6.00</td>
<td>11.76</td>
<td>5.46</td>
</tr>
<tr>
<td>Overall writing quality</td>
<td>16.83</td>
<td>2.64</td>
<td>16.02</td>
<td>3.22</td>
</tr>
<tr>
<td>Historical thinking rubric</td>
<td>4.50</td>
<td>2.67</td>
<td>5.28</td>
<td>2.70</td>
</tr>
<tr>
<td>Essay length</td>
<td>177.33</td>
<td>64.43</td>
<td>170.87</td>
<td>75.69</td>
</tr>
</tbody>
</table>
Transformations

A one-way analysis of variance test was conducted to evaluate the relation between the two genres and the number of transformations made when using documents to compose essays. There were no statistically significant differences on the total number of transformations $F(1, 101) = .059, MSE = 1.40, p = .81$ in students’ essays.

The results of a second $2 \times 2$ analysis of variance also indicated that there was no significant relation between level of content knowledge and the overall number of transformations students produced in their essays, $F(1, 101) = 3.87, MSE = 92.537, p = .05$. Moreover, there was no main effect for writing task, $F(1, 101) = 0.59, MSE = 1.40, p = .81$, nor an interaction effect between level of content knowledge and writing task $F(1, 101) = .006, MSE = .131, p = .94$. Thus, students’ level of content knowledge did not influence the total number of transformations they produced.

However, we performed a $2 \times 4$ between-subjects multivariate analysis of variance on the two writing genres and types of transformations in students’ essays (elaborations, paraphrases, additions, and misconceptions). The Wilks’ criterion indicated that this measure was statistically significantly affected by writing genre, Wilks’ $\lambda = .79, F(1, 99) = 6.28, p = .00$. There was a statistically significant main effect for elaborations, $F(1, 99) = 8.11, MSE = 18.25, p = .01$; and paraphrases, $F(1, 99) = 12.18, MSE = 83.16, p = .00$, but not for additions, or misconceptions. These results indicate that students included more elaborations in their arguments ($M = 3.44$) than in summaries ($M = 2.59$), while students included more paraphrases ($M = 4.47$) in summaries than in arguments ($M = 2.65$).

Writing Performance

We conducted two separate one-way analyses of variance to examine the relation between writing genre: argument and summary and the quality and length of students’ essays. In addition, we performed a $2 \times 4$ between-subjects multivariate analysis of variance to examine the relation between writing genre and the four components of historical thinking: substantiation, perspective, contextualization, and overall score. There was no statistically significant main effect for writing genre on writing quality ($p = .84$), or for the number of words students’ produced in their essays ($p = .91$). Similarly, results from the multivariate analysis of variance test indicated no difference in combined dependent variables, Wilks’ $\lambda = .76, F(1, 99) = 0.762, p = .56$, nor an effect for any of the elements of historical thinking ($p = .86$ for substantiation, $p = .20$ for perspective taking, and $p = .98$ for contextualization). Thus, these findings suggested that genre alone did not influence students’ writing performance.

Findings Related to Level of Content Knowledge

A second series of $2$ (level of content knowledge) $\times 2$ (writing genre) analysis of variance design was used to evaluate the relation between the same four variables and historical thinking, writing quality and length of students’ essays. There were no significant findings related to students’ level of content knowledge (low and high) before the investigation and genre with respect to historical thinking or overall writing quality. However, students’ level of content knowledge did affect the number of words they produced $F(1, 101) = 5.46, MSE = 24147.39, p = .02$. 


Students with high content knowledge wrote considerably longer essays than did students with low content knowledge \((M = 184.0\) and \(M = 152.4, \text{ respectively})\). Yet, there was no main effect for writing genre or an interaction between content knowledge and writing genre. Taken together, these findings suggest that genre and level of content knowledge minimally affect the extent that students can produce content on an assigned topic.

**Disability Status and Writing**

We made a final comparison of writing among students in the general education classroom and students with disabilities. Importantly, this was not done for the reading comprehension measure because the document sets were read to students with disabilities. We conducted three separate one-way analyses of variance to examine the relation between students with and without disabilities and the three writing measures: historical thinking, quality, and length of students’ essays. There was no statistically significant main effect for historical thinking \((p = .65)\), writing quality \((p = .96)\), or for the number of words students’ produced in their essays \((p = .58)\). These findings suggested that, after providing students with disabilities appropriate accommodations, they were able to write essays that were comparable to those written by their same-age peers in regards to historical thinking, quality, and length.

**DISCUSSION**

This study was designed to explore the effects of genre and level of content knowledge on students’ comprehension of and ability to demonstrate historical thinking with primary and secondary documents. Whereas professional historians write with either narrative or argumentative forms, we chose to compare writing summaries to arguments, in part because findings from educational research on genre with novices suggests benefits for summary over narrative (Wiley & Voss, 1999), because history teachers tend to privilege summary over narrative writing tasks, and because historians use an interpretive form of narrative that is absent from most secondary history classrooms. We were also interested in learning how differences in students’ level of content knowledge might influence their ability to think historically, on the basis of findings from prior research that such differences might substantively impact students’ performance in disciplinary comprehension and production measures. Last, we wondered how students who arguably struggle most with learning—that is, those with disabilities—would perform relative to peers who were not identified with academic struggles.

It is interesting to note that the results from this study indicate that genre, in itself, had modest effects on students’ ability to read and write from adapted primary source documents. In fact, our main finding was that students’ level of content knowledge, which was determined after learning about the historical topic, in combination with prompt genre, was most relevant. This was true with respect to students’ literal comprehension, as those with limited content knowledge about the topic who wrote summaries as well as students with adequate content knowledge who wrote arguments performed well on a multiple-choice reading comprehension measure. Moreover, while genre and content knowledge did not affect the total number of transformations that students made while reasoning with documents, students who wrote arguments included more elaborations, or explained ideas more often in their essays than students who wrote summaries. In addition,
students who wrote summaries included more paraphrases, copying more ideas than students who wrote arguments. These results align with prior disciplinary comprehension research that suggests students may demonstrate more effortful processing of information when writing an argument, which, in turn, may lead to a better understanding of the text (Stahl et al., 1996; Wiley & Voss, 1999).

In contrast, assigning students to compose arguments or summaries resulted in essentially the same degree of historical thinking (when examining their ability to engage in substantiation, contextualization, and perspective taking), their overall writing quality (when rating their focus, elaboration of content, organization, style, and writing conventions) and when considering the length of students’ papers. Moreover, factoring in the degree to which students knew about the historical topic was only relevant when considering their ability to elaborate ideas, as students who knew more about the historical topic wrote longer essays than students with limited knowledge—knowing more did not affect students’ writing quality or level of historical thinking. These results, while somewhat surprising, may point to the need for more research on the genre that students are asked to use when engaged in historical reading and writing tasks in school.

Third, the results from this study support the contention that students who are identified with disabilities could perform at levels commensurate with their general education peers, when given appropriate learning accommodations. In this study, students from these two groups differed on initial measures of reading and writing, as evidenced from their standardized test scores. However, we did not find differences between these groups on any of our reading or writing outcomes. Although it is possible that having only a small percentage of students with disabilities in our sample influenced these results, it also seems possible that the reading accommodation successfully mitigated the negative effects of these students’ disability. Reading the documents to students a few days before they were asked to reason with and write from them appeared to allow them to perform in ways that matched students who did not have identified learning challenges.

**Limitations**

We found a significant teacher effect (i.e., the individual who taught each class) on the historical writing outcome, $F(3, 97) = 12.52, p = .00$, meaning that despite using a common approach to teach students about the Gulf of Tonkin Incident and Resolution, teachers were not equally effective in providing background information. Examination of the post hoc tests indicated that the advanced placement (AP) teacher was most effective, which may have been due to the fact that he gave more input in creating the lesson. On the other hand, the importance of this problem may be less pronounced because each of the main findings regarding genre and influence of content knowledge stand, when analyses are repeated without the most effective teacher and students in the data set. A second limitation is the finding that that one lesson on historical context was not sufficient for providing all students with an adequate understanding of the Gulf of Tonkin Incident and Resolution.

**Conclusions**

The results from this study indicate that asking students to consider both sides of a contentious event and to provide evidence from primary sources in their essay may be more important than
genre in influencing high school students’ ability to reason historically. These findings align with results from prior work at this age level (Monte-Sano & De La Paz, 2012) that prompts that ask student to reconcile historical controversies are beneficial for promoting sourcing, corroboration, and contextualization. Further research is warranted to explore differences between our findings and those with older students (cf. Le Bigot & Rouet 2007; Wiley & Voss, 1998) on advantages in using argumentative prompts, yet our results suggest that teachers may consider crafting historical reading and writing tasks that elicit summary, as long as students are asked to engage in disciplinary tasks such as determining historical significance or grappling with controversies. We also believe the current findings are encouraging for students who struggle most with literacy, as prior research indicates argumentative writing can place a greater cognitive load on students than writing summaries. Last, our findings also affirm that history teachers must find robust ways to equip students with requisite content knowledge about historical controversies for them to be able to benefit most from disciplinary tasks such as reading and writing from primary source documents.

AUTHOR NOTES

Susan De La Paz is an associate professor in the College of Education at the University of Maryland. She has spent more than 15 years creating and validating writing curricula, first in English classes, and more recently within the discipline of history, that help teachers meet the needs of students with LD, struggling learners, and proficient and advanced learners in middle and high school classrooms. Her research interests primarily focus on designing interventions that advance adolescents’ disciplinary thinking and on teacher professional development. Daniel R. Wissinger is an assistant professor in the Department of Special Education and Clinical Services at Indiana University of Pennsylvania. He worked for nearly a decade as a Special Education teacher before earning his Ph.D. from the University of Maryland in 2012. Research interests include disciplinary writing, and writing interventions to support students with disabilities and those most at-risk for academic failure.

REFERENCES


August 5, 1964

Last night I announced to the American people that the North Vietnamese regime had conducted further deliberate attacks against U.S. naval vessel. The latest actions of the North Vietnamese regime have given a new and grave turn to the already serious situation in Southeast Asia. In recent months, these actions have become steadily more threatening.

The issue [at hand] is the future of Southeast Asia as a whole. A threat to any nation in that region is a threat . . . to us. This is not just a jungle war, but a struggle for freedom on every front. The North Vietnamese regime is conducting a campaign of subversion [rebellion], which includes the training, and supply of personnel and arms for the conduct of guerrilla warfare in South Vietnam. I should now ask the Congress to join in affirming the national determination that all such attacks will be met, and that the United States will continue in its basic policy of assisting the free nations of the area to defend their freedom.
Document 2: Statement by Senator Wayne Morse

Head Note: Senator Wayne Morse was one of the two Senators in Congress to vote against the Gulf of Tonkin Resolution.

August 7, 1964

Our government has no right to send American boys to their death in any battlefield without a declaration of war. And no war has been declared in Southeast Asia . . . until a war is declared it is unconstitutional to send American boys to their death in South Vietnam. I don’t know why we think, just because we are mighty, that we have the right to try to substitute might for right.

Since when do we have to back our President . . . when the president is proposing an unconstitutional act? I want to warn him I’m not giving him a blank check. This doesn’t mean that the president can go ahead and send additional troops over there without consulting us [Congress] . . . I most respectfully said that’s just nonsense. I have complete faith in the ability of the American people to follow the facts if you’ll give them. My charge against my government is we’re not giving the American people the facts.

Source: Excerpt adapted from Senator Wayne Morse’s interview with Phillip Babich from the National Radio Project on August 7, 1964.

Grade 8–9
Lexile measure 980L
Word Count 179

**Head Note:** On August 8, 1964 the New York Times reported on the passing of the Gulf of Tonkin Resolution.

Washington, Aug. 7 – The House of Representatives and the Senate approved today the resolution requested by President Johnson to strengthen his hand in dealing with Communist aggression in Southeast Asia. After a 40-minute debate, the House passed the resolution; 416 to 0. Shortly afterward the Senate approved it, 88 to 2. President Johnson said the Congressional action was “a demonstration to all the world of the unity of all Americans.”

The debates in both houses made clear, however, that the near-unanimous vote did not reflect a union of opinion... Congressman Nelson stated that our continuing policy is to limit our role to the provision of aid, training assistance, and military advice. It is the sense of Congress that, except when provoked... we should continue to attempt to avoid a direct military involvement in the Southeast Asia conflict.”

**Source:** Excerpt adapted from the New York Times Archive, August 8, 1964.

<table>
<thead>
<tr>
<th>Grade</th>
<th>8–9</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lexile measure</td>
<td>1020L</td>
</tr>
<tr>
<td>Word Count</td>
<td>155</td>
</tr>
</tbody>
</table>