UC Irvine
UC Irvine Electronic Theses and Dissertations

Title

Examining the features of students' source-based argument writing in history, epistemology, and the relations between them

Permalink

https://escholarship.org/uc/item/5z10p201

Author

Steiss, Jacob

Publication Date

2022

Copyright Information

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA,
IRVINE

Examining the features of students' source-based argument writing in history, epistemology, and

the relations between them

DISSERTATION

submitted in partial satisfaction of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

in Education

by

Jacob Steiss

Dissertation Committee Members:
Professor Emerita Carol Booth Olson, Chair
Assistant Professor Brandy Gatlin-Nash
Warner Professor Steve Graham
Associate Professor Penelope Collins

2022

Table of Contents

List of Figures

List of Tables

Acknowledgements

Vita

Jacob Steiss

EDUCATION

Expected 2022   **Ph.D., Education;** University of California, Irvine

May 2017   **M.Ed.;** University of Notre Dame

May 2014   **B.A., English and Sociology;** Fordham University

ACADEMIC EMPLOYMENT

Summer 2019-Present   **Leadership Team Member,** WRITE Center, University of California Irvine

Fall 2018-Fall 2019   **Community Research Fellow,** University of California Irvine

Summer 2017 and 2018   **Clinical Instructor,** University of Notre Dame

PUBLICATIONS

Steiss, J., Krishnan, J., Kim, YS., & Olson, CB. (2022). Dimensions of Text-Based Analytical Writing of Secondary Students. *Assessing Writing, 51,* 100600.

Steiss, J. (2020). Dismantling winning stories: lessons from applying critical literature pedagogy to the odyssey. *Journal of Adolescent & Adult Literacy,63*(4), 433-441.

Manuscripts in preparation/submitted for review

Steiss, J. Gatlin-Nash, B., & Collins C. Teaching about Dialect, Voice, and Race in To Kill a Mockingbird; The Affordances of a Critical Language -Literature Approach. *English Teaching: Practice & Critiqu*e. Submitted for review.

Olson, CB., Maamuujav, U. **Steiss, J.**, & Chung, H. Examining the Impact of a Cognitive Strategies Approach to Teaching Arguments of Literary Analysis on the Writing of Mainstreamed English Learners in Secondary School. *Written Communication*. Submitted for review.

CONFERENCE PRESENTATIONS

Steiss, J. & Krishnan, J. (2022, April 21-26). "Assessing the Dimensions of Source-Based Analytical Writing." In Graham, S. (Chair), *Multiple Lenses for Understanding Source-Based Analytical Writing Development*. Symposium conducted at the AERA Annual Meeting. San Diego, CA, United States.

Chung, H., **Steiss, J.**, & Olson, C.B. (2022, April 21-26) "A Cognitive-Strategies Approach to Argument Writing: Professional Development for Teachers of Secondary ELL Teachers." In Ferretti, R. (Chair), *Writing Disciplinary Arguments: Theoretical and Instructional Considerations.* Structured Poster Session conducted at the AERA Annual Meeting. San Diego, CA, United States.

Steiss, J. & Krishnan, J. (2021, December 2-6). "Research-based Strategies for Teaching Argument Writing in History" In Pearson, D. (Chair) Research-based Strategies for Teaching Argument Writing: Cognitive, Sociocognitive/Sociocultural, and Design-Based Perspectives. Symposium conducted at LRA Annual Conference. Atlanta, GA, United States.

Steiss, J. & Krishnan, J. (2021, May 7-8). *Practical Design and Assessment Strategies for Improving Source-Based Argument Writing (Even at a Distance)*. Teaching History Conference. Hosted by the History Project at UC Davis through an online platform.

Steiss, J., Aubele, J., Gilbertson, N., Martin, D., & Fellowes, S. (2021, May 7-8) *Preparing Students to Navigate Controversial and Contested Histories in the Past and Present: The Benefits of Integrating Media Literacy and Historical Thinking.* Teaching History Conference. Hosted by the History Project at UC Davis through an online platform.

Steiss, J. & Krishnan, J. (2021, April 9-12*). "Tools to Expose": How Students Can Disrupt Canonical Texts Through Critical Literacy and Analytical* Writing [Roundtable Session]. AERA Annual Meeting. Online.

Steiss, J. (2020, February 21-23). *Exploring Critical Approaches to The Odyssey and Other Commonly Taught High School Texts* [Conference Presentation]. California Association of Teachers of English. Los Angeles, CA, United States.

Chung, H., **Steiss, J.,** & Olson, C.B. (2022, April 21-26). "Cognitive-Strategies Approach to Argument Writing: Professional Development for Teachers of Secondary ELL Teachers." In Ferretti, R. (Chair) *Writing Disciplinary Arguments: Theoretical and Instructional Considerations*. Structured Poster Session conducted at AERA Annual Meeting AERA Annual Meeting. San Diego, CA, United States

Steiss, J. (2020, Apr 17 - 21) *Understanding Practitioners' Perspectives in Research-Practice Partnerships* [Structured Poster Session]. AERA Annual Meeting San Francisco, CA http://tinyurl.com/yywlb7a3 (Conference Canceled)

Teacher Professional Development Presentations

Aubele, J., **Steiss J.**, Gilbertson, N., & Martin, D. "Helping students become responsible media consumers, curators, and producers." Three-part Workshop delivered virtually on February 9, 2022, February 16, 2022, and February 23, 2022. Sponsored by *UC Santa Cruz History & Civics Project* and *UC Irvine History Project*

Steiss, J. & Maamuujav, U. "The Path to Writing Proficiency for English Learners." Workshop delivered virtually on January 15, 2022 to 7 school districts participating in the Pathway Project.

Gilbertson, N., Martin, D., **Steiss, J.**, & Aubele, J. "Building Students' Media Literacy for Civic Engagement Fall Series." Three-part Workshop delivered virtually on Sep 17, 2020, Oct 1, 2020, and Oct 15, 2020. Sponsored by *UC Santa Cruz History & Civics Project* and *UC Irvine History Project*

Gilbertson, N., Martin, D., **Steiss, J.**, & Aubele, J. "Building Students' Media Literacy for Civic Engagement Spring Series." Three-part Workshop delivered virtually on February 9, 2021, February 16, 2021, and February 23, 2021. Sponsored by *UC Santa Cruz History & Civics Project* and *UC Irvine History Project*

Steiss, J. & Rhee, J. (2017). "Increasing relevancy, engagement, and dialogue." Presentation delivered at *For the Greater Glory of God: Innovation and Collaboration for the 21st Century*, Archdiocese of Los Angeles All Teachers Professional Development Day.

Steiss, J., Rhee, J., Lopez, N., & Mann, T. (2016). "Promoting cognitive engagement for 21st century learners." Presentation delivered at *ACE Teaching Fellows Conference,* University of Notre Dame

GRANTS/AWARDS

Spring 2022	$5,000 - Project: *Developing Critical Literacy Across Disciplines: A Three-Part Workshop;* Funding provided by the UCI Teacher Academy
Spring 2022	$5,000 - Project: *Collaboratively Building Students' Media Literacy for Civic Engagement;* Funding provided by the UCI Teacher Academy
Fall 2021	$3,000 - Project: *Building teacher knowledge and efficacy for teaching media literacy and civic reasoning.* Funding provided by the Digital Literacy Accelerator: Awarded by Department of Education, Office of Educational Technology
Fall 2020	$1,000 - Project: *Establishing The Media Literacy Collaborative Teacher Research Group.* Funding provided by the Center for Research on Teacher Development and Professional Practice
May 2017	$6,500 Eli Segal Education Award for Service, Americorp
February 2016	$500 College Football Playoff Foundation Grant
September 2013	$19,000 Archbishop Hughes Scholarship for Promising English Majors

TEACHING EXPERIENCE

Winter 2021	**Teaching Assistant**, University of California, Irvine

EDUC 348 Equity and Exceptional Learners; Instructor of Record: Dr. Brandy Gatlin-Nash

Summer 2017 and 2018   **Clinical Instructor**, University of Notre Dame, *Alliance for Catholic Education*

Guest Lecture

November 2021   EDUC 305 Learning to Learn from Teaching

Examining student work for what matters most

February 2021   EDUC 348 Equity and Exceptional Learners

Designing Culturally and Linguistically Responsive Lessons using UDL

October 2019   EDUC 346 Reading and Writing in MS/HS Classrooms

Practical Strategies for Promoting Discussion across Content Areas

November 2019   EDUC 346 Reading and Writing in MS/HS Classrooms

Making Discussion Equitable in Diverse Classrooms

Secondary Teaching Experience

2017-2018   **English Teacher,** Xaverian Brothers High School

2015-2017   **English and Social Studies Teacher,** CSHM High School

2014-2015   **Long-Term Substitute Teacher,** Inter-Lakes High School

PROFESSIONAL SERVICES

Committee Member - DECADE, First-year student mentor (2018-2021); University of California, Irvine

Panelist - Cross-Cohort Conversations (2018-2021); University of California, Irvine

Ad Hoc Reviewer - *Language, Speech, and Hearing Services in Schools;* Collaborative review with Doctoral Advisor, Brandy Gatlin, PhD.

Ad Hoc Reviewer - *Assessment in Education: Principles, Policy & Practice;* Collaborative review with Doctoral Advisor, Carol Booth Olson, PhD

Abstract of the Dissertation

Examining the features of students' source-based argument writing in history, epistemology, and the relations between them

by

Jacob Steiss

Doctor of Philosophy in Education

University of California, Irvine, 2022

Professor Emerita Carol Booth Olson, Chair

Developing students' source-based argument writing skills is a vital educational goal for the 21st-century information society. Consequently, researchers and educators continually seek ways to understand and improve students' capacities for constructing and advancing arguments based on multiple documents, texts, or sources in a range of subject areas in secondary education. To contribute to these efforts, this dissertation explores the relations between students' source-based argument writing and a factor increasingly seen as a predictor of how students reason and write with multiple sources—their epistemology—beliefs about knowledge, and how it is constructed. In three studies, I examined: 1) secondary students' source-based argument writing skills in history, 2) students' epistemologies in this discipline, and 3) the relations between student writing and epistemology. Findings from all three studies will help researchers and educators better understand students' source-based argument writing skills, their views about the nature and construction of knowledge, and how these are related.

Introduction

Source-based argument writing is an important and complex skill

National and state standards have increasingly emphasized reasoning with and writing arguments using multiple sources (Common Core State Standards Initiative, 2012; Levine, 2014; National Council for Social Studies, 2013). Makers of the CCSS have emphasized that argument literacy and reasoning with evidence are ''essential to both private deliberation and responsible citizenship in a democratic republic'' (National Governors Association Center for Best Practices & Council of Chief State School Officers, 2010, p. 3). The increasing emphasis on complex literacy skills in the standards reflects the demands of a 21st-century information society where students must critically evaluate and synthesize information across sources, provide coherent explanations to inquiry questions, and support claims with relevant evidence and analysis (Goldman et al., 2012; Rouet and Britt 2011). Indeed, many of the complex social, civic, and scientific questions that students face require making reasonable judgments and arguments based on multiple sources (Applebee & Langer, 2011; Bråten et al., 2011a).

Although multiple-source-based reasoning is developed and assessed in a variety of ways, source-based argument *writing* (SBAW) is seen as a particularly useful, appropriate, and flexible way to develop and assess student competencies and progress towards complex reasoning and writing skills across content areas in secondary schools.

To illustrate, Appendix A of the CCSS describes argument writing as having a particularly important place in secondary curricula:

> While all three text types are important, the Standards put particular emphasis on
> students' ability to write sound arguments on substantive topics and issues, as this ability
> is critical to college and career readiness. English and education professor Gerald Graff

(2003) writes that "argument literacy" is fundamental to being educated. The university is largely an "argument culture," Graff contends; therefore, K–12 schools should "teach the conflicts" so that students are adept at understanding and engaging in argument (both oral and written) when they enter college.

As students enter college or more fully participate in civic life, forming arguments is a crucial skill (CWPA, 2011).

Despite the importance of argumentation, students have few educational opportunities to read and construct written arguments based on multiple sources (Applebee & Langer, 2011; Hastings et al., 2012). Further, SBAW is an incredibly complex and challenging task that requires the strategic coordination of multiple, overlapping cognitive and social processes (Goldman & Scardamalia, 2013; Rouet & Britt 2011).

Recent NAEP results indicate secondary students are generally challenged with writing, and a wide body of research documents the complex challenges prompted by SBAW, especially its length, complexity, and the use of multiple sources (Goldman & Scardamalia, 2013; Rogers et al., 2013). Secondary students struggle to evaluate and select evidence that is reliable and relevant to claims (Goldman & Scardamalia, 2013; McGrew et al., 2018; Rouet and Britt 2011), integrate multiple sources (List et al., 2019), present both sides of an argument or issue (Anmarkrud et al., 2014), and use source information to predict, interpret, or evaluate a document's content (Bråten et al., 2011b; Britt & Aglinskas, 2002). To help students develop the multifaceted literacy skills needed for SBAW, secondary schools must integrate source-based writing instruction across content areas (not just in ELA classrooms) to prepare students for the digital age and 21st-century information society.

The importance of argument literacy across content areas

To ensure opportunities to practice source-based argument writing across content areas, the Common Core State Standards (CCSS) and the College, Career, and Civic Life Framework (C3) emphasize the development of literacy skills students need for college and career success through authentic disciplinary inquiry (CCSSI, 2012; Cowgill II, & Waring, 2017; Levine, 2014; NCSS, 2013). In other words, students need to learn to argue in a discipline as they learn *how* people in that discipline argue. Proponents of disciplinary literacy assert there are subject-specific ways of reading and writing that are needed to perform the distinct tasks of a discipline (Moje, 2008; Shanahan & Shanahan, 2008, 2012). The types of texts read, the practices for reading and writing, and what counts as valid knowledge claims, evidence, and sound argumentation vary from discipline to discipline (Langer 2011; Moje, 2008).

A historian thinks of data differently than a scientist or literary critic and also interprets this data with different tools. Thus, disciplinary writing presents different rhetorical issues and different problems for the writer to solve. History teachers engage students in primary source analysis, crafting historical questions, and constructing meaning from facts with no clear answer (Bickford, 2010; Breakstone et al., 2013; Monte-Sano, 2010; VanSledright, 2002; Wineburg, 1991).

Further, views about knowledge in disciplines also vary (Cribb et al., 2018; Goldman et al., 2016; Shanahan & Shanahan, 2012). Those practicing the disciplinary norms of history view interpretations of events as tentative, unconfirmed, and liable to be disproved with countervailing evidence (Bain, 2006; Monte-Sano, 2010; Monte-Sano & De La Pas, 2012). Practically, this has myriad implications that must be considered for understanding writing. For one, acknowledging and determining the validity of counterclaims is a crucial part of source-based arguments that must be addressed in writing evaluation and instruction (Bain 2006; Goldman et al., 2016;

Monte-Sano, 2010, 2012; Monte-Sano & De La Paz, 2012; Wineburg, 1991). Further, it implies that to write well in a discipline one needs to *think* well in a discipline; this implication is a core premise of this dissertation.

Still, there are fundamental literacy practices that span disciplines. Argument writing in history requires knowledge and skills that are both general and discipline-specific (Hillocks, 1995; Monte-Sano, 2010). Argument frameworks, like Toulmin's (1958), outline components of arguments that apply across disciplines (Monte Sano & De La Paz, 2012). As mentioned previously, National reports and research indicate the challenges students face exist across disciplines, including task interpretation, presenting both sides of an argument or issue, and supporting arguments with evidence (Anmarkrud et al., 2014; Du & List, 2020; Goldman & Scardamalia, 2013; List et al., 2019; NCSS, 2013; Olson et al., 2012, 2017)

While it is unclear to what extent argument writing and historical writing are separate (Monte-Sano, 2010), recent research has affirmed the impact of approaches that integrate general literacy development alongside disciplinary literacy (De La Paz et al., 2017; Graham & Perin, 2007; NCES, 2012; MacArthur et al., 2019; Song & Ferretti, 2013). In history classrooms, for example, students need general declarative, procedural, and metacognitive knowledge implicated in argument writing (Flower & Hayes, 1981; Hayes, 2012; Hillocks, 1995, 2011; McCutchen, 2006), such as the use of claims, data, warrants, and counterarguments (Toulmin, 1958), but also disciplinary knowledge like placing evidence into its historical context when making arguments (Monte-Sano, 2010). Other disciplinary skills include analyzing textual evidence from primary and secondary documents, determining central ideas from multiple sources, determining an author's point of view, analyzing relationships between sources, and constructing evidence-based arguments (Breakstone et al., 2013; NCSS, 2013).

By integrating SBAW skills and disciplinary literacy practices (e.g.. engaging in historical inquiry using primary sources) students can develop general and specific literacy skills jointly (Lee & Swan, 2013; De La Paz et al., 2017). Attending to all these literacy skills means students produce knowledge in writing that is both "general and discipline-specific" (Monte-Sano, 2010, p. 543) Such an approach was adopted in this dissertation, utilizing a broad body of research on what supports students' writing development while also giving attention to the disciplinary considerations of students in situated environments, specifically history classrooms, and more specifically, SBAW tasks with multiple sources.

Given that current research indicates high school students and college freshmen use relatively unsophisticated approaches when it comes to SBAW writing within and across disciplines (Braasch et al., 2009; Kuiper et al., 2005; NCES, 2015; Wineburg, 1991), this dissertation contributes to the field by describing what students' SBAW looks like in history in and across grade levels. Because history is a discipline that lends itself well to the development of general and disciplinary argumentation, findings are relevant to researchers and educators concerned with general and content-specific literacy practices. In addition to describing key components of SBAW in history, this dissertation also examines a key attribute increasingly acknowledged as influencing source-based reasoning and writing—epistemology—one's beliefs about knowledge and how it is constructed.

The role of epistemology in source-based argument writing

Beliefs about the nature of knowledge and knowing have been shown to affect information processing, evaluation, comprehension, argumentation, source integration, and writing across a variety of students and content areas (Bråten et al., 2011a; Britt & Rouet, 2012; Ferguson & Bråten, 2013; Richter & Maier, 2017; Wiley et al., 2020). Further, epistemology

predicts subcomponents of source-based argument writing in history, such as attending to the source of information to make inferences about its relevance and reliability (Barzilai & Eshet-Alkalai, 2015; Barzilai et al., 2015; Bråten et al., 2011b; Bråten et al., 2014; Britt & Aglinskas, 2002; Strømsø et al., 2013).

Epistemology influences the task models students adopt as they answer an inquiry question (Bråten et al., 2011a; Britt et al., 2012). For example, a student viewing history as a static set of facts might adopt the goal in a writing task to find and succinctly present the right answer to the historical question, while ignoring any conflicting evidence. A student with a more adaptive epistemology, that is, one in line with norms for knowledge construction in a discipline, might adopt a different task model. This student sees knowledge as complex and tentative. Consequently, as they read, think, and write they weigh multiple potential responses to the historical question before arguing why one explanation has more evidentiary support than another. Because such a student might produce writing that differs substantially, it is worthwhile to investigate the extent to which epistemology contributes to discrete components of writing (e.g. text integration and sourcing moves) in a discipline as thinking well in that discipline is directly linked to writing well.

Students' beliefs about knowledge have generally been approached from either a dimensional (Bråten & Strømsø, 2009; Voss et al., 1998; Wiley et al., 2020) or developmental view (Barzilai & Eshet-Alkalai, 2015; Kuhn & Weinstock, 2002; Maggioni et al., 2009). The dimensional view examines beliefs about the nature of knowledge (whether it is simple or complex) and beliefs about how we know (knowledge is certain or tentative) as distinct dimensions. These two dimensions may be manifest in distinct ways in a history classroom. For example, a student who sees knowledge as complex may endorse multiple reasons as to why the

Montgomery Bus Boycott succeeded (i.e. the boycott succeeded due to Rosa Parks' decision to stay seated *and* the actions of those who organized the carpool once the boycott of public buses began). This same student might also see historical interpretations as tentative and revise their understanding of the event when new sources include alternative explanations (e.g. Jo Ann Robinson distributing hundreds of flyers in the community to spread word about the boycott).

A developmental view sees students progressing through stages from an absolutist (the facts are the facts), to a multiplist (many different interpretations exist), before becoming an evaluativist who sees knowledge claims as verifiable through evidence-based interpretation (Kuhn, 2001). This final stage matches the disciplinary practices of historians and supports the acquisition of advanced literacy skills (Alexander et al., 2011; Goldman et al., 2016; Monte-Sano, 2010; VanSledright, & Maggioni, 2016). Absolutist views of knowledge are seen as naive and multiplist views as somewhat more sophisticated. Similarly, from the dimensional perspective, a view of knowledge as complex and tentative is seen as adaptive whereas a view of knowledge as simple and certain is seen as naive and maladaptive to source-based reasoning tasks.

Yet, many researchers advance a view of students as seeing history as certain and preferring single cohesive narratives (Voss, et al., 1995; Voss et al., 1998). Such a view conflicts with adaptive epistemologies that see history as an evidence-based interpretation of events that would lead to better thinking, corroboration, and more sophisticated arguments about historical topics (Bain, 2005; Monte-Sano, 2008; Wiley et al., 2020). Therefore, close attention to the writing of students, their epistemologies, and relations between the two will be informative for educators and researchers in the field.

Understanding epistemology and its relation to writing also matter for SBAW's relevance

to 21st-century literacies and problem-solving. Individuals with more adaptive epistemologies not only construct better arguments but can reason with multiple conflicting sources of evidence and weigh multiple accounts of events (Barzilai & Chinn 2020; Goldman et al., 2012; Kuhn, 2019), traits increasingly important for civic participation in an information society in which truth and accuracy are increasingly elusive (Kavanagh & Rich, 2018; Kendeou et al., 2019). This civic goal underscores the purpose of history education more broadly, that historical reasoning, thinking about the causes, consequences, and significance of historical events, matters for students' civic reasoning today. Given the present focus on argument writing, a key civic skill, findings from this dissertation have implications for civic discourse. In other words, the quality of arguments about civic or social issues in society is related to student's reasoning skills as well as their beliefs about knowledge, how it is constructed, and how we come to know and agree on what is true as a society (Barzilai & Chinn, 2020).

Given the complexities and importance of source-based argument writing, this dissertation contributes to the following: 1) understanding the features and quality of students' SBAW across grade levels and subgroups of students, 2) understanding the nature of students' epistemologies, and 3) understanding the relations between the student SBAW and epistemology. By accounting for the relations between epistemology and components of student writing, the three studies produce models of students' SBAW development across grade levels that can lead to improved instruction that targets the traits, skills, and contexts needed for flourishing in the "intertextual reality" characteristic of the 21st-century information society (Bråten et al., 2011a, p. 49). For example, if epistemology is found to reliably predict students' skills in evidence use in a written argument, this suggests educators could productively devote time to addressing students' beliefs about history to improve key writing skills. Therefore,

findings are relevant to researchers and educators seeking to improve students' *reasoning* and *writing* in source-based inquiry tasks.

The studies of the dissertation

The first study contributes to emerging understandings of students' source-based argument writing skills in history. The study was guided by the following research question: *What are the features of secondary students' source-based argument writing in history?* To address this question, a research team developed an analytic framework that measures features of students' SBAW in history. The framework was applied to writing samples from students participating in the field trial year of a randomized control trial aiming to improve students' SBAW skills in two school districts. Findings included descriptive statistics of student writing as measured by the analytic framework which measures discrete components of student writing. Findings also showed differences in writing between grade level and EL status. Confirmatory Factor Analysis (CFA) and Structural Regression (SR) were used to examine the dimensions of SBAW in history and the relative contributions of each dimension to the holistic score.

In the second study, I answered the following research question: *What are the features of secondary students' epistemologies in history?* I answered this question by developing two scales to measure students' epistemology in history—one adopting a dimensional approach and one adopting a developmental approach. Data from the epistemology scale and CFA were used to examine the dimensions of students' epistemologies in history. These dimensions were [intended to be] used as latent factors in the next study which described the relations between writing and epistemology.

The following question guided my third study: *What are the relations between students'*

9

epistemologies and their source-based argument writing? Given the growing evidence that epistemology influences student writing (Bråten et al., 2011a; Britt & Rouet, 2012; Ferguson & Bråten, 2013; Richter & Maier, 2017; Wiley et al., 2020), I used structural regression to predict the effect of epistemology on the dimensions of writing confirmed in Study 1. Findings indicate to what extent epistemology influences writing and what components of writing are most affected.

Significance of the studies

Findings from all three studies help researchers and educators better understand students' writing capacities, their beliefs about knowledge in a discipline, and how their beliefs about knowledge relate to or affect their writing in a discipline. The range of students (grade 6-12) and analytic approach contribute to the understanding of when students (and important subsets of students such as English Learners) develop adaptive writing skills and beliefs about knowledge.

Findings related to epistemology and its relation to student writing are relevant to researchers seeking to understand epistemology as a potential moderator of SBAW quality. Further, given the paucity of historical thinking skills in student writing (e.g. sourcing) and their importance to the 21st-century information society, the studies show when key skills are developed across grade levels and among certain subgroups of students such as English Learners. Whereas many studies investigating the relations between epistemology and components of writing use dichotomous indicators or only holistic scores, the more analytic approach to measurement employed in the studies was more sensitive to variation in writing skills and which subcomponents are most related to epistemology.

The skills and epistemologies examined in this dissertation have relevance beyond the secondary [history] classroom as well. Many of the argument writing skills in history, such as

using source information to evaluate the relevance and reliability of a claim, and many of the indicators of an adaptive epistemology, such as seeing truth as the resolution of competing interpretations through analysis of evidence, are foundational to civic reasoning in a 21st-century information society. The inability of citizens to reason with multiple conflicting sources of evidence or attend to alternative perspectives is disruptive to civic participation and must be addressed through approaches that attend to students' information literacy skills as well as their views about how truth and knowledge are formed (Barzilai & Chinn 2020; Kavanagh & Rich, 2018).

Theoretical framework

A wide range of sociocognitive and sociocultural research tells us that myriad individual and contextual factors influence the process of source-based argument writing with multiple sources. Given this, across all three studies, I drew on the MD-TRACE (Multiple-Document Task-based Relevance Assessment and Content Extraction) model which describes how individuals engage in multiple document learning tasks (Rouet & Britt, 2011). It outlines a multi-step process individuals use when engaging in tasks with multiple sources (e.g. answering a historical question using primary sources) and notes that differences in how individuals perceive tasks, differences in context, and differences in external and internal resources available to individuals (e.g. epistemology or epistemic beliefs) will influence task engagement and subsequent products (e.g. writing).

In their review of instruction promoting multiple source use, Barzilai and colleagues (2018) identify MD-TRACE as playing a major role in instruction using multiple sources across history, science, and language arts in secondary schools. The MD-TRACE model developed from the refinement of the Documents model (Perfetti et al., 1999; Britt et al., 1999). A wide

body of research shows the model provides broad insights into how individuals reason with multiple sources, documents, or texts (Bråten et al., 2011a; Bråten et al., 2014). Evidence that supports MD-TRACE or a documents model framework comes from research in history (e.g., Britt & Aglinskas, 2002; Nokes et al., 2007; Rouet et al, 1996; Wolfe & Goldman, 2005) and other content areas (Bråten et al., 2011a).

Broadly, the model posits that the way individuals approach a task with multiple sources is influenced by a combination of external and internal resources—how an individual interprets and creates a mental model for the task, their prior knowledge, their self-regulation, their understanding of source content, their understanding of the relations between sources, environmental cues, the schema individuals currently have for solving similar types of problems, and their beliefs about knowledge (Rouet et al., 2017). To illustrate, task interpretation matters because individuals will disengage from a task once they feel the standards of the task have been met. Consequently, students who think answers to historical questions are simple and straightforward and do not necessarily need to be supported with evidence from multiple sources will produce a written argument that is brief and underdeveloped. Therefore, one's beliefs about knowledge, epistemology, can determine goals for engagement, the standards for text production, and the final product of students thinking and production.

To illustrate further, research has documented that individuals who differ in their beliefs about the amount of intertextual integration that must occur to answer a historical question will produce different writing (Limon & Mason, 2002; Wiley et al., 2020). As integrating evidence from multiple sources and addressing counterarguments is important for a strong response to a document-based question in history class, understanding and attending to students' beliefs about knowledge is important for understanding and improving their writing.

Another prominent source-based reasoning model identified by Barzilai and colleagues (2018) is Wineburg's (1991) model of historical thinking, which illustrates the cognitive processes historians engage in when analyzing and evaluating primary sources, namely sourcing, contextualization, and corroboration. Sourcing, a process used across disciplines and integral in the MD-TRACE framework, means attending to information about the source of a document/text/media to assess its relevance and reliability. Corroboration refers to the intertextual process of comparing information across sources (e.g. whether two different sources agree on some key point) and is also a key cognitive process in multiple document-based tasks across disciplines. Contextualization as a practice is arguably the most discipline-specific and involves situating sources within their temporal and spatial context to better understand their perspectives.

Wineburg's work to identify the key cognitive strategies involved in historical thinking has been widely used to study reading, cognition, and writing in history. For the present studies, MD-TRACE and historical thinking are not mutually exclusive, but rather complementary. The complementarity of the two theories is echoed by other research as well. For example, proponents of MD-TRACE often attend to individuals' sourcing capabilities because evaluating and using information about who, what, and why a document was produced can influence how an individual constructs a model of sources and their relations to each other (Barzilai et al., 2015; Bråten et al., 2014; Strømsø et al., 2013). Sourcing is often an essential component for the satisfactory completion of multiple-document inquiry tasks, like resolving disparate accounts of historical narratives (Bråten et al., 2017; Braasch & Bråten, 2017). Corroboration, comparing evidence across sources, is another way that one might meet the demands of an inquiry task (as they interpret these demands). In this way, both frameworks are compatible, with historical

thinking identifying discipline-specific tools individuals utilize within the MD-TRACE model.

Though the MD-TRACE model has also been recently adapted to the RESOLV (REading as problem SOLVing) model to more fully account for the role of the task, discipline, and social context in students' reading processes (Rouet et al., 2017), this model mostly focuses on the *reading* process of individuals. Similar to MD-TRACE, the model posits that how students perceive the demands of a disciplinary task will influence how they approach and carry out the tasks (Wiley et al., 2020). This can occur across disciplines (e.g. reading for universal meaning in language arts vs. reading to understand the past in history), and within disciplines (e.g. writing an essay to get an "A" vs. interpreting a historical document to better understand the past). Although the RESOLVE and MD-TRACE models are similar, the MD-TRACE model was centered in the present studies due to RESOLV's emphasis on the reading process.

Researchers adopting the MD-TRACE model explicitly emphasize that individual differences in epistemology can influence reasoning and literacy outcomes, such as writing performance, analytical thinking, sourcing, and the representation of other viewpoints in argumentation (Abendroth & Richter, 2020; Barzilai et al., 2015; Barzilai & Eshet-Alkalai, 2015; Bråten et al., 2014; Christensen-Branum et al., 2019; Du & List, 2020; Kammerer et al., 2015; McCrudden & Sparks, 2014; Strømsø et al., 2013; Wiley et al., 2020; Wolfe, 2012). Because epistemology influences the task models readers, thinkers, and writers adopt as well as their thinking with specific sources of information, it is worth identifying students' epistemologies as well as potential moderators of their reading, writing, and thinking.

For example, students taking an evaluativist approach to a multiple source-based inquiry task begin the task recognizing that there will be multiple competing interpretations to the essential question (Barzilai & Zohar, 2012; Mason & Boscolo, 2004). They also are more likely

to engage in specific reasoning processes, like recognizing texts are constructed in specific times and places. Students with more naïve epistemologies may not attend to source information to resolve discrepancies and consequently fail to answer a historical question adequately (Bråten et al., 2011b; Britt & Aglinskas, 2002; Wineburg, 1991; Wiley et al., 2020).

Epistemology may also play a key role in writing development due to its metacognitive role in knowledge production (Barzilai & Zohar 2014). Because metacognition plays a key role in reasoning or arguing using multiple sources (Moshman, 2018; Shraw & Moshman, 1995), helping students determine and control how cognitive effort is spent and regulated most effectively will be a key component of an intervention to improve source-based writing (Graham et al., 2015, 2016). For example, certain individuals may ignore conflicts, rather than engage in effortful resolution, due to an evaluative stance that reflects a naive epistemology (List & Alexander, 2017; Richter & Maier, 2017; Stadtler & Bromme, 2014). As described previously, if a student believes that the Montgomery Bus Boycott succeeded largely due to the role of Rosa Parks and this student also believes knowledge in history is simple and certain, they will be less likely to seek more information from sources that posits different actors and actions as leading to the success of the movement. Their epistemology preserves a stable and coherent view of the historical issue.

Positioning epistemology in the MD-TRACE model can allow for a fuller description of the factors leading to high-quality writing within a given discipline. Given the interaction between students writing and epistemology, I now review relevant literature related to students' SBAW in history, students' epistemology, and the relations between the two. I then outline my three studies, how they fit within this research base, and the unique contributions they will make to the field.

Background and literature review

Components of source-based argument writing of secondary students

Because writing is a complex composite of many interrelated skills, researchers and educators approach assessment and instruction from both holistic and *analytical* approaches (MacArthur et al., 2019; Olinghouse et al., 2015). In holistic assessment, one can determine a student's overall writing proficiency (Charney, 1984). Through an analytical approach, one can measure proficiency across discrete but related skills and then tailor instruction to students' needs. This approach can offer more precise insights into student performance by assigning different scores components of student writing (Bacha, 2001). For example, the 6+1 trait system is used widely in the U.S. and other countries (Gansle, et al., 2004; NREL, 2011) to assign student scores for the following categories: idea development, organization/structure, word choice, sentence fluency, voice, presentation, and conventions.

Research measuring specific components of writing considers the age of the students, the genre of the writing, and the writing task before determining what components of writing may plausibly be observed in the writing (Kim & Graham, 2021; Mo & Troia, 2017; Wagner et al., 2011). For example, studies with elementary-aged students often focus on substantive quality, organization, productivity, syntactic complexity, and spelling and writing conventions (Kim et al., 2014; Wagner et al., 2011). Text-based genres, including literary analysis and argument writing, require skills like the integration of textual evidence and commentary and are, therefore, more complex when compared to non-text-based genres (Olson et al., 2020; Schleppegrell, 2004). The additional elements also require different approaches to assessment and instruction. As the present studies take an analytical approach to measuring secondary students' SBAW in history, a review of the components of writing in this genre and by students of this age is

16

warranted.

As mentioned previously, national standards like CCSS and C3 standards require that students display many complex skills in their argument writing: "cite textual evidence from primary and secondary documents, determine central ideas from primary and secondary documents, determine author's point of view, analyze charts and graphs, distinguish between fact and opinion, analyze relationships between primary and secondary documents, and construct evidence-based arguments (CCSSI, 2013; NCSS, 2013). The C3 standards also emphasize the development of disciplinary skills (Lee & Swan, 2013). In this way, SBAW in history comprises general literacy skills, such as advancing a claim, but also discipline-specific ones, such as identifying the context and perspective of primary source documents to determine their reliability (Breakstone et al., 2013; Wineburg, 1991). In what follows, I describe components of writing that can be conceived of as general literacy skills and disciplinary skills, with both being essential to writing in history, but disciplinary literacy skills not necessarily being relevant to source-based writing in other disciplines (e.g. science and ELA). These components also function as hypothesized latent factors in the CFA model tested in Study 1.

Presentation of Ideas and Structure

Broadly, the presentation of content and ideas is frequently referenced when assessing writing quality (NREL, 2011; National Writing Project, 2010) and is described as distinct from language-based features of writing (Kim et al., 2014; Troia et al., 2019; Wilson et al., 2017). The presentation of a strong claim and addressing the writing prompt is seen as distinct from other components of writing (e.g. evidence use) in multiple subjects (Correnti, et al., 2013; Wang, et al., 2018). In history, the presentation of ideas can be referred to as substantiation, which describes how well the writing offers explanations in support of a claim (De La Paz, 2017 et al.;

17

Monte-Sano, 2010; Wiley et al., 2020).

Further, the presentation of ideas also subsumes traits like persuasiveness and the structure and presentation of argument structure (Monte-Sano, 2010). Because the presentation of ideas includes what is said as well as how clearly it is communicated or *presented* (Steiss et al., 2022; Wagner et al., 2011) the ideas of writing and organization can be thought of as a specific dimension of writing. While some popular writing rubrics may see structure and ideas as separate components, this has not been validated to my knowledge.

Evidence Use

Researchers have observed the complexity and variation in evidence use across developmental levels in writing (Correnti et al., 2020; O'Hallaron, 2014; Wang et al., 2018) and have seen these skills as a distinct component of students' source-based writing in history (De La Paz, et al., 2012; Monte-Sano, 2010, 2011; Pessoa et al., 2019). In history, skills related to evidence use include selecting the best evidence, attributing evidence to the source material, and reasoning with evidence to support claims (Monte-Sano, 2010; Reisman et al., 2019; Brimsek, & Hollywood, 2019; Van Drie & van Boxtel, 2008). The challenge of integrating and using evidence to support claims is well documented (De La Paz et al., 2012; Monte-Sano, 2010; Wiley & Voss, 1999), though less certain claims can be made related to whether evidence use is a district or dissociable dimension of writing in history. The present studies will address this question.

Historical Thinking

Historical thinking encompasses several skills related to reasoning with historical evidence, including analyzing evidence through historical moves such as sourcing, contextualizing, and corroborating to assess the validity of different historical interpretations or

perspectives (Monte-Sano 2011; Wineburg, 1991). Some research suggests historical thinking is a separate dissociable dimension of writing quality in history, though one correlated with overall quality (De la Paz et al., 2017; Monte-Sano, 2012; Pessoa et al., 2019). Researchers have identified key features of historical thinking in student writing, including sourcing, contextualization, perspective recognition, and addressing counterarguments (De La Paz et al., 2017; Du & List, 2020; Monte-Sano 2010, 2011, 2012; Wiley et al., 2020). These skills are notably difficult for students to display proficiently across disciplines (Bråten et al, 2011b; Britt & Aglinskas, 2002; Lund et al., 2019; Wineburg, 1991).

Historical thinking is seen as crucial to argument writing using multiple sources in history. Using source information, or sourcing, allows readers to organize sources and their content when creating a representation of a historical event and can help determine which claims are best supported in response to a historical question (Bråten et al., 2011b; Britt & Aglinskas, 2002). Further, because historical thinking helps students understand the intention of the author, the context of writing, the position/perspectives of sources, and the reliability of the evidence, it is necessary to make an accurate and persuasive argument in response to historical questions (Gottlieb & Wineburg, 2012; Paxton, 2002).

Language Use

Writing quality indicators related to language use (e.g. conventions, diction, syntactic variety) are also distinct components in commonly used writing rubrics. Empirically, studies have found specific language features to uniquely explain variance in student writing (MacArthur et al., 2019; Troia et al., 2019). Studies assessing the dimensionality of writing have found syntactic complexity and conventions to be related but dissociable constructs of writing, albeit with younger students (Kim et al., 2014, Kim et al., 2015). Linguistic features of writing

have also been shown to vary by genre and prompt (Crossley, 2020; Kyle & Crossley, 2016).

Therefore, language use is a possible distinct dimension in the SBAW of secondary students in history.

Approaches to measuring students' beliefs about knowledge

Beliefs about the nature of knowledge and knowing, epistemic beliefs, have been shown to affect information processing, evaluation, comprehension, argumentation, and performance in literacy tasks across a variety of disciplines (Bråten et al., 2011a; Ferguson & Bråten, 2013; Richter & Maier, 2017; Wiley et al., 2020) Though the inclusion of measures of epistemology in history is relatively novel (Maggioni et al., 2009; VanSledright & Maggioni, 2016; Wiley et al., 2020), epistemology has been used as a variable to explain adolescent reasoning and literacy outcomes across wide contexts and subjects. Broadly, researchers interested in epistemology have taken either dimensional or developmental views of epistemology, both of which will be examined in the present studies.

A dimensional view of epistemology

Preliminary work measuring individuals' epistemology (Hofer, 2001; Hofer & Pintrich, 1997; Schommer 1990, 1994) has advanced multidimensional views of epistemology. Recent empirical work has validated and argued for the use of four dimensions to identify an individual epistemology for descriptive or predictive applications. Two dimensions reflect an individual's views about the *nature* of knowledge (how simple/complex knowledge and how certain/complex knowledge is) and two dimensions reflect an individual's views about the nature of how one comes to know (what is considered a good *source* of knowledge and how one *justifies* a knowledge claim). Figure 1 shows how one can conceive of an individual falling somewhere on a spectrum in each dimension of epistemology. For example, the top half of the figure shows

how one's beliefs about the *nature* of knowledge can be relatively naive, believing knowledge is simple and certain, or they can be relatively sophisticated, believing knowledge is complex is tentative. The bottom half of the figure shows different beliefs about how knowledge is *constructed*. One may see an individual's intuition as a reliable or even preferred source of knowledge, whereas someone else may see knowledge as primarily formed through evidence-based reasoning and distrust knowledge claims based on intuition.

Figure 1.
Dimensional View of Epistemology
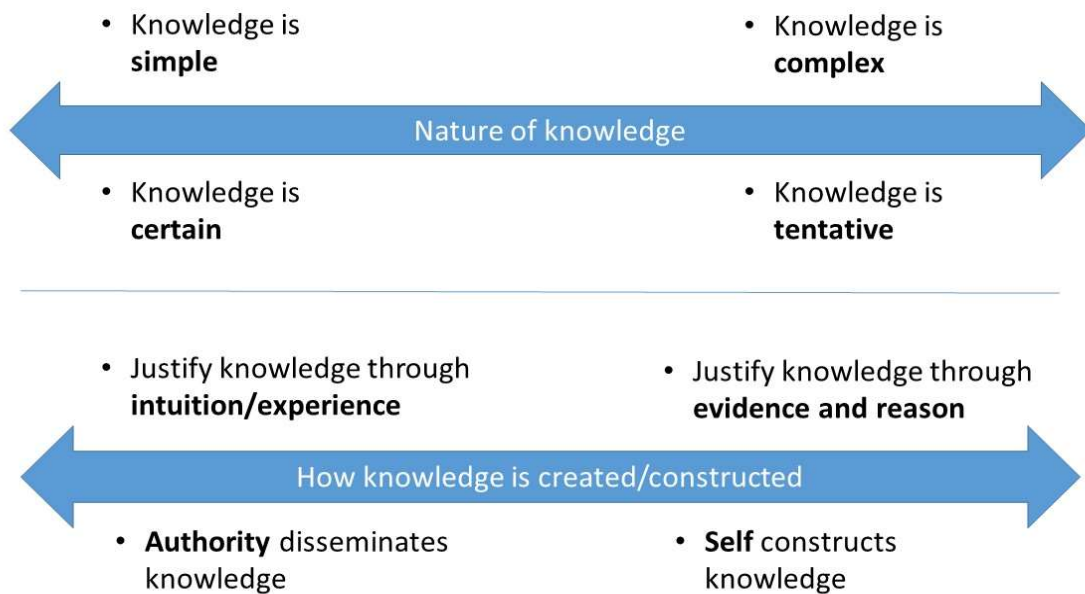
Researchers have outlined numerous ways in which one's position on this spectrum matters for their reasoning, comprehension, evaluation, argumentation, and writing (Barzilai & Eshet-Alkalai, 2015; Bråten, et al., 2008; Rukavina & Daneman, 1996; Sinatra et al., 2003). In history, for example, viewing knowledge as complex leads to better historical reasoning and constructing cause/effect explanations (Wile, et al., 2020). See Table 1 for more information about each dimension and a practical example of how variability in specific dimensions predicts different thinking processes or behaviors.

21

Table 1.

Prominent Dimensions of Epistemology

Dimension of epistemology	Practical implication
Simple/complex dimension: views of knowledge; from isolated facts to complex interrelated concepts	Belief in knowledge as simple facts makes individuals less motivated to synthesize information from multiple documents into a coherent framework (Strømsø et al., 2008) Viewing arguments as facts, not constructions, with multiple sides, makes one less likely to endorse other views (Wolfe, 2012; Christensen-Branum et al., 2019) Braten and Strømsø (2006) argue a complex view of knowledge is good for building coherence across multiple conflicting texts
Certain/tentative dimension: view of knowledge absolute and static to tentative and evolving	Believing knowledge about an issue is complex leads to less use of sources that simplify an issue (e.g. a newspaper article that depicts a one-sided view of an issue) (Strømsø et al., 2013) Kardash & Scholes, (1996) argue a view of knowledge as tentative and complex leads to more open-minded writing Awareness of the constructed nature of knowledge helps improve the integration of multiple online sources (Barzilai & Zohar, 2012)
Source of knowledge dimension: from originates outside self, with external authority to actively constructed in persons socially	Belief in knowledge from an authority, vs. constructed from self has mixed results in reading comprehension and text interpretations (Bromme 2005; Bråten et al., 2008)
Justification of knowing dimension: from justifying claims in observation, intuition, and personal experience to rules of inquiry and evaluation	Braten and Strømsø (2010) argue that beliefs that knowledge claims (e.g. what led to the success of a historical movement) should be based on inquiry and evaluation of sources are more adaptive than beliefs that common sense or first-hand experience should be the source of knowledge claims This dimension is key in source evaluation as a higher-order moderating process (Bråten et al., 2014) as individuals will be more inclined to assess source material and integrate evidence from multiple sources

In three of the four dimensions, researchers see more sophisticated beliefs as generally more adaptive—it is better to think of knowledge as complex and tentative (rather than simple and certain) and to seek to justify claims through inquiry and evaluation. To illustrate, a common task in historical interpretations is to reconcile two conflicting accounts of a historical event. Given conflicting sources and asked, "Who is responsible for the violence that occurred in the "Boston Massacre?", students who believe knowledge is complex and tentative will likely comprehend both sides of the event better, assess the reliability of sources (instead of ignoring sources that disagree with prior views), and write explanations that account for multiple perspectives. Similarly, believing that claims should be justified through evidence and interpretation will result in more integration of evidence across sources.

 In the source of knowledge dimension, however, research has shown mixed results as to what end of the spectrum is more adaptive. Mason and colleagues (2006) found that beliefs in personal meaning construction positively affected holistic text interpretation, including understanding main ideas, and the production of text inferences. Buehl and Alexander's (2005) work produced contradictory findings as students with high beliefs in authority as the source of knowledge performed quite well regarding text-based learning. These results suggest that opposite ends of the spectrum are more or less adaptive depending on the context, sources, and task (Bråten et al., 2014). For example, it may be adaptive for an individual with low knowledge about climate change to trust authority, but in challenging an interpretation of a literary interpretation it may be more adaptive to trust one's own ability to construct knowledge.

Ferguson and Bråten (2013) found that Norweigan 10th graders with high knowledge of an issue, a lack of belief in personal justification, and a high-level belief in justification by multiple sources performed best on multiple-text comprehension measures. One implication

could be that this dimension of epistemology is not stable or reliable. Another implication is that one must flexibly shift between trusting oneself as a constructor of knowledge and trusting authority based on topic complexity and one's own beliefs and knowledge. For example, given one's low knowledge about how vaccines are developed, it is reasonable to trust claims about vaccine safety from reliable experts and not construct knowledge about this topic from one's intuition, experiences, or "research."

One must also consider the discipline in which one is reasoning as beliefs can be more or less adaptive in different disciplines. Given this, the first two dimensions, Simple/Complex and Certain/Tentative were used to study epistemology in history because, in this discipline, the belief that knowledge is constructed, tentative, and complex is needed to spur effortful evaluation or elaboration with conflicting information (Richter & Maier, 2017). I measured these two dimensions in Study 2 of this dissertation. The fourth dimension, how one believes knowledge claims should be justified, was also addressed in the second study as evidentiary reasoning is a key component of history as a discipline (Goldman et al., 2016). Wiley and colleagues (2020) took an approach to more directly measure beliefs in the value of evidence and reasoning to resolve interpretations in history that is adopted in the present studies by including justification through the integration of evidence (Integration) as a dimension of epistemology in history. Finally, I did not attend to the third dimension, sources of knowledge, given issues in its measurement in previous research, a lack of clarity about its relation to reasoning and writing outcomes, and a lack of research showing which end of the spectrum is adaptive in history (e.g. it is unclear the extent to which trusting a source of authority, like a textbook, is adaptive to productive reasoning and writing).

A developmental view of epistemology

Researchers studying epistemology have also taken a developmental approach, measuring individuals' progress from naive to more sophisticated epistemologies as represented by three categories: absolutist, multiplist, and evaluativist (Kuhn, 1999, 2001; Kuhn & Weinstock, 2002). Individuals progress in their epistemology, first perceiving knowledge as a static and objective reality, then an arena of multiple subjective opinions, before moving to a final stage which integrates the previous two stages. An individual becomes an evaluativist when they "[understand] knowledge as judgment based on evaluation in a framework of alternatives and evidence and, accordingly, subject to change" (Kuhn, 2019 p. 28; Greene et al., 2016).

This approach to measurement overlaps with the dimensional approach as it attends to how individuals conceptualize knowledge and how individuals believe one comes to know. However, as opposed to measuring dimensions separately, a developmental view posits that within a domain (i.e. history) an individual's epistemology is unidimensional and can be categorized as absolutist, multiplist, or evaluativist. Each of these categories represents an individual's view toward knowledge and informs their approach to solving information problems in a given domain.

The following table summarizes major differences between the three categories, which are also seen as developing stages--an individual moves from being an absolutist, to being a multiplist, before becoming an evaluativist as they reach epistemological maturity.

Table 2.
Naive to Mature Epistemologies

Developmental stage	Practical implication
Absolutist: Believes knowledge is concrete, stable, and a set of indisputable facts; there is one correct view of the world	An absolutist who does not believe in climate change may dismiss contradictory claims as false and in need of correction; They wouldn't see their view about climate change to be falsifiable even with more evidence

Multiplist: Believes multiple truths can exist in the world; each subjective view of an issue can be right in its own way	Increasing awareness of multiple perspectives can support the move towards epistemological maturity but can fall short of reaching an objective understanding of important issues; To say climate change believers and non-believers are both correct is implausible and potentially catastrophic.
Evaluativist: Believes knowledge can be produced through rational inquiry; there are sometimes competing theories about a topic, but these can be resolved by using evidence and reason	Knowledge is complex and tentative, but claims about a topic are falsifiable; Therefore, we can know if climate change is real and true and we should use valid epistemic processes to discover the truth

Note. Table created referencing research across different domains (Barzilai & Weinstock, 2015; Barzilai & Zohar, 2012; Greene et al., 2016; Kuhn, 2019; Mason & Boscolo, 2004).

The developmental view is rooted in argumentation and discourse studies (Kuhn, 2019) and has been used to understand why certain individuals write more effective and balanced arguments. For example, Mason and Scirica (2006) found that 8th grade students who were evaluativists wrote more effective arguments and included more counterarguments and rebuttals compared to their multiplist or absolutist peers. Similarly, moving towards an evaluativist epistemology has been linked to higher reading comprehension and understanding of complex issues (Kuhn, 1999; Kuhn & Winestock, 2002). While there are cognitive processes in reconciling divergent views, one's epistemological maturity can also moderate this process. For example, one needs to first believe there are multiple views on an issue before being compelled to formulate a counterargument. One then needs to believe knowledge claims can be disproved (evaluativism) before moving to do so. In this way, epistemology can be seen as a type of epistemic metacognition that influences how tasks are carried out and how arguments are structured (Barzilai & Zohar, 2014); Epistemology then is a key process that determines how one works and carries out multiple source-based tasks within the MD-TRACE model.

Relations between epistemology and source-based argument writing

Theoretical and empirical findings support further investigation of the relations between epistemology and SBAW. In their meta-analytic review, Greene and colleagues (2018) concluded that epistemology can predict academic performance across a variety of domains and tasks. Such an influence may exert itself in several ways. For one, epistemology has been conceptualized as a type of metacognition, or epistemic thinking, monitoring one's thinking about knowledge claims and how to justify them throughout the work of reading, integrating, and writing with multiple sources (Barzilai & Zohar, 2012; Bromme et al., 2010; Hofer, 2001; Mason & Bromme, 2010). Most cognition, including epistemic thinking, needs to be seen as valuable for one to justify the effort (Kuhn, 2019). For example, even if students have demonstrated they can integrate information from multiple sources, they might not do so in different settings (e.g. evaluating claims on Twitter vs. synthesizing sources in an academic essay) because they do not see such cognition as relevant to the task at hand or valuable in such a setting (Barzilai & Zohar, 2014; Chinn and Buckland, 2012).

Epistemic thinking also affects information processing, evaluation, and comprehension (Bråten et al., 2011; Britt & Rouet, 2012). Therefore, in tasks that require students to integrate multiple sources of information to answer a question, for example, an argument about the most important causes of a historical event, epistemology can be a crucial predictor of second-order evaluation or validation processes (McCrudden, & Barnes, 2016). Practically, students with a belief that knowledge is simple and certain may not attend to information that conflicts with what they previously know or have read about a historical event. This lack of deep processing can also be tied to a low standard for coherence typical of students with more simple views of knowledge or with weak epistemic thinking (Du & List, 2020).

Studies have also linked students' epistemology with their sourcing activities both while reading and writing (Barzilai et al., 2015; Barzilai & Eshet-Alkalai, 2015; Bråten et al., 2014; Wiley et al., 2020). These findings support the theory that epistemology leads to more effortful evaluation needed to resolve conflicting information (Richter & Maier, 2017) and epistemology's role in the integration of source content, attribution of evidence to source material, and incorporating different perspectives or viewpoints, especially those that conflict with prior knowledge or beliefs (Britt, Perfetti, Sandak, & Rouet, 1999; List & Alexander, 2017; Perfetti, Rouet, & Britt, 1999; Taber & Lodge, 2006).

Epistemology has also predicted better integration of information about a topic across texts (Rukavina & Daneman, 1996; Stromso & Braten, 2006) and better multiple text understanding (Bråten et al.,, 2008), both of which are crucial for making a coherent and organized argument in a source-based inquiry task. Further, an adaptive epistemology, that knowledge is complex, tentative, and evolving is seen as predicting attention to multiple sources and viewpoints (Barzilai and Eshet-Alkalai 2015; Bråten et al. 2011). As integrating evidence from multiple sources and addressing counterarguments are key indicators of quality in SBAW, epistemology is a key component leading to quality writing.

Domain-specificity: Epistemology in argument writing in history

Epistemology must also be considered within the context of SBAW tasks (Kuhn, 1999). An adaptive epistemology is determined by the demands of the task and what is considered an appropriate construction of knowledge in a discipline (Britt et al., 2013; McCrudden, & Sparks, 2014; Rouet and Britt 2011; Rouet, Britt, & Durik, 2017). Therefore, epistemology can be both domain-general and domain-specific (Muis et al., 2006), with most researchers electing to study epistemology using domain-specific instruments.

Recent concerns with epistemology in history can be linked to an emphasis on disciplinary differences in knowledge construction. Such differences influence practices in a discipline (Alexander et al., 2011; Goldman et al., 2016; VanSledright, & Maggioni, 2016). Beliefs about what history is, how knowledge is created in history, and how knowledge claims are justified influence thinking and growth in historical understanding (VanSledright, & Maggioni, 2016). To measure epistemology in history, one needs to consider the object of study, the subject of history as a discipline, and the practices and standards for producing knowledge in history.

For the present studies, it was important to operationalize an adaptive epistemology as one that would produce accurate and valid knowledge about history. Two recent trends in secondary history instruction are relevant to this dissertation: 1) the increased use of SBAW tasks and 2) adopting an alignment with disciplinary literacy practices to improve literacy and content knowledge. Disciplinary literacy refers to the specific ways experts in a discipline think, read, write, and communicate (Goldman et al., 2016; Shanahan & Shanahan, 2012) and includes their processes and products of communication as well as their beliefs about knowledge and how it is constructed (i.e. epistemology). A brief explanation of the epistemological stance in the discipline of history and its relation to student writing is warranted.

Regarding the objects of study, history teachers often use textbooks at the expense of historical artifacts (i.e. primary sources). The dominance of textbook-based instruction and an absence of source-based inquiry tasks are increasingly considered problematic. That current students see history as a body of facts (Goldman et al., 2016; VanSledright, 2002) is seen as tied to the predominance of textbook-based instruction (Bain, 2006), which is not well suited to teaching the complex disciplinary reasoning needed for source-based inquiry tasks. Instead of

teaching history as a "fixed story" (Monte-Sano, 2010, p. 1063), other researchers and educators

argue it is important to have an epistemology in history that predisposes one to inquiry and

evidentiary thinking (Ashby, Gordon, & Lee, 2005; Bain, 2005).

As most textbooks reflect a belief that knowledge is static, simple, and rests in authority,

researchers argue that history classrooms should turn to document or source-based inquiry which

more accurately reflects history as a competing set of interpretations supported by evidentiary

thinking (De La Paz et al., 2017; Monte-Sano, 2010; Nokes, 2013; Young & Leinhardt, 1998).

Such practices are informed by a more adaptive epistemology that emphasizes a view of history

as evidence-based interpretations that can always be revised, corroboration across sources, and

the weighing of different viewpoints before coming to a reasoned conclusion (Bain, 2005, 2006;

Goldman et al., 2016; Monte-Sano, 2010; Reisman, 2012). In such a classroom, students are

more able to do the work of history, learn content knowledge, and practice disciplinary literacy

skills that are seen as relevant to 21st reasoning skills: forming a coherent interpretation of an

event from multiple documents, advancing strong arguments with supporting evidence and

reasoning and using evidence to justifying knowledge claims (Ashby, Gordon, & Lee, 2005;

Goldman & Scardamalia, 2013; Lee & Shelmit, 2003;).

Seeing history as an evidence-based interpretation of events has been seen as adaptive to

writing improvement (Monte-Sano, 2008) and a recent study by Wiley and colleagues (2020)

suggests student epistemology was predictive of major differences in the writing of secondary

students, including key historical thinking moves like sourcing and contextualization. This is in

line with the suggestion of the MD-TRACE model that internal resources and perceptions of the

task will influence the use of disciplinary practices like inquiry, corroboration, and forming

historical interpretations from evidence (Maggioni et al., 2009). Increasing source-based

instruction can develop students' more adaptive epistemologies. At the same time, an adaptive epistemology may be necessary to participate in such disciplinary practices competently. If a student sees history as a stable and simple body of knowledge enshrined in a textbook, one would not expect that student to be well-prepared to resolve competing interpretations of a historical event through conflicting sources of evidence.

Further, in other subjects, epistemic beliefs are independent predictors of performance by secondary students (Greene et al., 2018) The present studies' contributions to this growing base of knowledge are, in part, aided by a more comprehensive measurement of student writing and the use of both dimensional and developmental views of epistemology as potential moderators of student writing performance. Though some efforts have been made to measure epistemology in history, they have focused on teachers (Maggioni et al., 2009), have demonstrated low consistency scores for the proposed scales, and have been unclear as to whether epistemology is best viewed as uni- or multidimensional (Maggioni et al., 2009; Stoel et al., 2017; Wiley et al, 2020).

Study 1: What are the features of secondary students' source-based argument writing in history?

Purpose

Given the need to better understand the features of secondary students' argument writing in history, the first study examined source-based argument writing (SBAW) of a sample of students in grades 6, 7, 8, 10, 11, and 12 (History is not taught in grade 9 in this state). The study was guided by the following research question: *What are the features of secondary students' source-based argument writing in history?* The study examined differences between grade levels and between students designated as English Learners across discrete components of writing. The study also examined the dimensions of SBAW in history and their relations to holistic scores. Dimensions of SBAW writing in history were used in Study 3 which examined the relations between epistemology and writing.

Given the broad challenges students face when writing arguments with multiple sources, I hypothesized that students' overall writing performance would be similar on average across discrete features of writing such as the presentation of ideas, structure, and language use, but students would especially struggle with features of writing related to historical thinking and evidence use (namely, sourcing, contextualization, and corroboration, attribution, and presenting reasoning to support a claim).

I also expected differences between grade levels and between students of different language statuses to be significant across all components of writing, with older students performing better than younger students and students designated as EO/IFEP, and RFEP performing better than students designated as EL. I expected that among high school students in the sample, components of writing related to their evidence use and historical thinking would

have more variation than components only related to their language skills (e.g. sentence fluency, diction, syntactic variety and style), which aligns with research showing the development of academic language takes time (Olson et al., 2017)

Using Confirmatory Factor Analysis, I also hypothesized that SBAW in history would best be represented by a five-factor solution—Presentation of Ideas, Structure, Language Use, Evidence Use, and Historical Thinking— with all five factors, except Historical Thinking, being moderately related to each other. Finally, I hypothesized that all dimensions of SBAW in history would predict holistic scores.

Method

Participants

Participants in the first study included 207 secondary students from two urban school districts in the southwest United States. Students were part of a field trial for a writing intervention aimed at improving secondary students' SBAW through improved teacher knowledge and instruction. In this pilot study, 24 teachers participated in professional development to improve students' SBAW in history across grades 6, 7, 8, 10, 11, and 12. At each grade level, three teachers in district A and one teacher in district B (4 teachers per grade level) selected one classroom to participate as a focal class to test the efficacy of the intervention to improve student writing outcomes.

Across the four teachers at each grade level, stratified random sampling was used to select 9 students per teacher, or 36 students per grade. Stratified random sampling procedures first blocked students by sex and then by English language status (as determined by the California Department of Education's definition of English Learner), to ensure an adequate number of students designated as IFEP (Initially Fluent English Proficient), RFEP (Reclassified

33

Fluent English Proficient), EO (English Only), and EL (English Learner) were included in the study. In each class, I randomly selected 1 student designated as IFEP or RFEP per sex, 2 students designated as EO per sex, 1 student designated as EL per sex, and 1 additional student of any language designation. One 8th grade teacher participating in the pilot study changed teaching assignments before the school year began and did not have her grade 8 students participate in the intervention. Therefore, the stratified random sampling of 9 students over *23 classrooms* resulted in a sample of 207 students used in the present study.

In the sample, approximately 16% of the students were designated as ELs, 27% were designated as RFEP, and 57% were designated as EO/IFEP. These percentages are similar to district-wide percentages for each district. In the sample, 48% of the students were female. The districts did not provide the racial/ethnic composition of the students or the free or reduced lunch (FRPL) status of the students in the study. However, at a district level, 73% of students received FRPL in District A, 70% received FRPL in District B, and over 70% of students in each district were Hispanic/Latinx.

Measures

Source-based Argument Writing Task

To examine the features of students' SBAW in history, I examined student performance on a baseline writing measure. During the field trial, students were randomly selected at the classroom level to write to one of two text-based analytical writing prompts, which were administered across two, 50-minute class periods. Each prompt asked students to read four sources about a historical topic and to write an argument of causal analysis. Students wrote arguments in response to one of the following questions: 1) *How did the Montgomery Bus Boycott succeed?* (BB, n=99), or 2) H*ow did the Delano Grape Strike and Boycott succeed?*

(GB, n=108). All students responded to the prompt using Google docs and typed their essays. All students were familiar with using Google docs and keyboards were provided for all students. An option to hear the sources read aloud was also provided to all students. See Appendix A for the complete version of each writing prompt.

The prompts were designed the previous year over multiple cycles of testing, analysis, and modification. Both prompts exhibit the writing skills emphasized in history classrooms as well as Common Core State Standards (CCSI, 2012; Goldman et al., 2016; Cribb et al., 2018; Martin et al., 2011). Both prompts use a mix of both primary and secondary sources and emphasize constructing interpretations of the past using evidence and reasoning (Bickford, 2010; Breakstone et al., 2013; Wineburg, 1991). As document-based questions require the use of different, even conflicting sources to understand the past, they reflect the processes of historical inquiry and disciplinary practices of historians (Rouet et al., 1996; Wiley et al., 2020; Wiley & Voss, 1996; Wineburg, 1991). Further, because historical inquiry with multiple sources is an act of constructing truth or knowledge in history, it is influenced by epistemology. Students with different views about the nature of knowledge and how it is constructed may engage in historical inquiry differently and create different task products (Cowgill II, & Waring, 2017). In the prompts, students can advance clear and direct causal explanations of the past or complex explanations with multiple causes. Therefore, the prompts provide an opportunity to exhibit wide variation in writing and are sensitive to detecting performance differences between students with distinct epistemologies.

The prompts also present distinct boundaries between sources (i.e., placing source text in boxes and on separate pages), make source information salient in each source with headnotes and footnotes in bold text, and modify sources according to researchers' guidelines (Britt &

Aglinskas, 2002; Britt et al., 2013; Lund et al., 2019; Wineburg & Martin, 2009). Modifications included eliminating extraneous vocabulary, modifying length, and eliminating irrelevant proper nouns. Lastly, the prompt, *How did the Montgomery Bus Boycott succeed?,* was adapted from a similar lesson created by the *Stanford History Education Group* (Stanford History Education Group, n.d.).

Holistic scoring

During the field trial, trained evaluators assigned holistic scores to student writing. A holistic scoring rubric was developed using extant measures for evaluating SBAW and writing in history and was shared with subject matter experts in the field to assess content validity. The rubric used a scale of 1 to 6, with a 1 indicating "No Evidence of Achievement" and a 6 indicating "Exceptional Achievement." The holistic rubric captured all criteria related to proficient SBAW in history: the presentation of a clear and compelling argument that addresses the requirements of the prompt; strong organization and coherence; the use of relevant and sufficient evidence; reasoning that connects the evidence to claims; the selection, integration, and attribution of relevant evidence from source materials; the presentation of and response to counterclaims; the quality of historical thinking related to the analysis of the source material and historical reasoning; and the use of sophisticated and appropriate academic language. See Appendix B for the complete rubric. 18 raters were trained in the use of the holistic rubric using anchor papers and identifying key features of papers scoring in the 6 different categories 1-6. All of the essays were double-scored to assess reliability. The average agreement within 1-point was 89% for the evaluators. Scores that disagreed by more than 1 point were scored by a third evaluator.

Analytic framework for source-based argument writing in history

While holistic scoring speaks to the overall quality of SBAW (Charney, 1984; Schipolowski & Böhme, 2016), trained coders also used an analytic framework to evaluate discrete components of students' writing (e.g. the quality of reasoning and how well the writing attributes evidence to sources). The analytic framework independently measures components of text-based analytical writing that, taken together, represent all the aspects of writing in the holistic score. This is similar to the use of an analytic rubric, but the items within analytic categories, such as Language Use and Evidence Use, for example, are broken down even further. To illustrate, evaluators using the NWP-AWC analytic rubric (National Writing Project, 2005, 2010) assign a score to the writing's "Content" by weighing multiple subcomponents of the writing's content, including the presentation of a thesis and the quality of the analysis of textual evidence. The analytic framework measures these criteria as separate items and assigns each item a unique score.

To create a reliable and valid framework, the research team first generated a list of items representing criteria used for the assessment of SBAW quality in history. Items were designed to measure all of the separate components of text-based analytical writing that, taken together, would represent overall writing quality in this genre. The research team drew on extant writing rubrics, research examining the quality of writing in history, and input from subject matter experts in generating items (Goldman et al., 2016; Monte-Sano, 2010, 2012, 2015; Monte-Sano & De La Paz, 2012; National Writing Project, 2005, 2010; NREL, 2011) Items were continuous (e.g. "How well does the writer present a clear and compelling) and were scored on a scale of 1-7; with a 1 indicating "not evident" and a 7 indicating "highly effective."

After the independent generation of items, the research team assessed the representativeness of items. The team then compared the full list of items with criteria outlined in

both holistic and analytic rubrics to determine whether the items taken together were a true composite of writing quality. Steps were taken to reduce items that were redundant or inapplicable. Next, the team applied prototypical analytic frameworks to a sample of student papers to better assess the usefulness of each item and determine what additional information was needed. After several iterations of testing, generating items, and reduction, the tool was shared with 8 subject matter experts in the field of secondary writing research who provided critique and feedback on the analytic framework, especially as it relates to content validity (Anastasi & Urbina, 1997).

Table 3 shows analytic framework items grouped by analytic categories that are reflected in analytic trait rubrics used to assess distinct components of SBAW across content areas. Confirmatory factor analysis was used to determine whether these categories represent the dimensions of SBAW in history. The table also shows interrater agreement rates, which are discussed in the following section.

Table 3.

Analytic Framework Items to Measure Writing Quality in History

Analytic Dimension	Item	Description	Inter-rater agreement
Ideas	Address	How well does the writing address all aspects of the prompt?	**100%**
	Pclaim	How well does the writing present a clear and compelling claim	**100%**
	Focus	How well is the writing focused on proving/substantiating an interpretable claim?	**93%**
Structure	Intro	How well does the writing present a skillful introduction with context, direction, and a clear claim	**89%**
	Body	How well does the body present a structure that enhances the central argument?	**93%**
	Conc	How well does the conclusion relate to claims made throughout and give the writing a sense of completeness?	**89%**
	Org	Writing has sound macro organization with a strong introduction, body, and conclusion	**96%**

Evidence Use	Evidence	How well does the student use source material as evidence? (Integrates sufficient evidence that is compelling and consistently relevant to claim(s))	**96%**
	Commentary	How does commentary interpret and use the textual evidence (to support a claim)?	**96%**
	Balance	How well does the writing balance purposeful summary, evidence, and commentary?	**96%**
	Attrib	How well does the writing attribute evidence to sources?	**96%**
Historical Thinking	Sourcing	To what extent does the writing use source material for sourcing? | Identifying and understanding the opinions, positioning, and bias of the author of a particular document; assessing credibility	**93%**
	Context	To what extent does the writing use contextualization? | Locating actors and actions (from source material) in their time and place; their social and historical contexts	**89%**
	Corrob	To what extent does the writing use corroboration? | Checking sources against each other to determine the validity of a claim	**96%**
	Counter	To what extent does the writing present and address alternative viewpoints/opposing perspectives?	**93%**
Language Use	Fluency	How well does the essay demonstrate sentence fluency and sentence flow?	**93%**
	Syntax	How well does the essay demonstrate syntactic variety and style?	**87%**
	Diction	How well does the essay demonstrate command of diction and word choice? (Discount language "borrowed" from source text)	**80%**
	Conventions	How well does the essay demonstrate control of language and standard grammar conventions including spelling, capitalization, and punctuation?	**100%**
	Tone	To what extent does the writing adjust language and use tone appropriate to purpose, audience, and task?	**100%**

Note: Raters used a 7-point scale; reported agreement rates within 1-point

Coding Team

To achieve high rates of interrater agreement, the research team took the following steps: 1) made individuals responsible for coding no more than five items in the analytic framework; 2) engaged in iterative cycles of coding to clearly define criteria used to code essays and improve reliability; and 3) generated a list of anchor papers (MacArthur et al., 2019) that described each

score (1-7) for every item in the framework. A description of each item can be found in Table 3. Coding teams were expected to reach a high degree of interrater agreement before proceeding to code independently. Two researchers oversaw each team's progress to ensure the content validity of their measurements.

15% of the essays (n= 32) were double-coded (Gallagher et al., 2017). Interrater agreement rates were calculated for each item of the framework. For all categorical items, agreement within a score point (on a 7-point scale) was considered acceptable (Bang, 2013; Gallagher et al., 2017; MacArthur et al., 2019) whereas exact agreement was necessary for dichotomous items. The average agreement within 1 point for all categorical items was 94% and all agreement rates were above 80%.

Analytic approach

Descriptive statistics

To answer the research question, *What are the features of secondary students' source-based argument writing in history?,* I examined descriptive statistics of student writing as measured by the analytic framework which intends to measure discrete components of student writing (e.g. presenting a counterargument) that may be differentially related to the overall quality of students' writing and dimensions of students' epistemology. Descriptive statistics were examined by grade level, with ANOVA used to determine significant differences by grade level across items of the analytic framework.

Confirmatory Factor Analysis and Structural Regression

Before engaging in Confirmatory Factor Analysis (CFA), items in the analytic framework were examined for their univariate and bivariate distributions and multivariate normality given the importance of certain distributional assumptions for Structural Equation Modeling (Kline,

2015). Given the infrequent observations of historical thinking moves like sourcing, contextualization, and corroboration in student writing, data transformations were considered at this stage before using these variables in CFA models. Bivariate correlations between all analytic items and holistic scores were also examined before conducting CFA.

Data from the analytic framework and CFA, using Mplus 8.4 (Muthén & Muthén, 2017), were used to examine the dimensions of SBAW in history. Four competing alternative confirmatory factor models shown in Figure 2 were fitted to the data from the analytic framework. The first model tested a unidimensional model, the baseline model, where SBAW in history is a single construct that captures all the items in the analytic framework (Figure 2a). The second model tested a multidimensional model, where Ideas/Structure, Evidence Use, Historical Thinking, and Language Use are dissociable, but related dimensions of writing quality. This model posits that the ideas and structure of an argument essay are too closely related to be dissociable constructs, as seen in recent research (Steiss et al., 2022).

The third model was similar to the second in that it tested the assumption of multidimensionality. However, this model (Figure 2c) proposed a correlated five-factor model (Figure 2b) where five factors—Presentation of Ideas, Structure, Language Use, Evidence Use, and Historical Thinking—are indicated by hypothesized item groupings described in the analytic framework. Finally, a fourth model was tested: a bifactor model (Figure 2d) with a general factor indicated by all the variables and specific factors indicated by the factor structure in Figure 2b or 2c (depending on the relative fit of the four and five-factor models). In a bifactor model (Gibbons & Hedeker, 1992), the general factor (overall writing quality) captures common variance across all the indicators while the specific factors, orthogonal to the general factor, help to explain variance that is not captured by the general factor. Model comparisons were conducted

41

using chi-square tests of difference, given that these models were nested, as well as other fit

indices (Hu & Bentler, 1999; Kline, 2015)

Figure 2.
Competing Models of SBAW in History

Figure 2a

Figure 2b

Figure 2c

Figure 2d
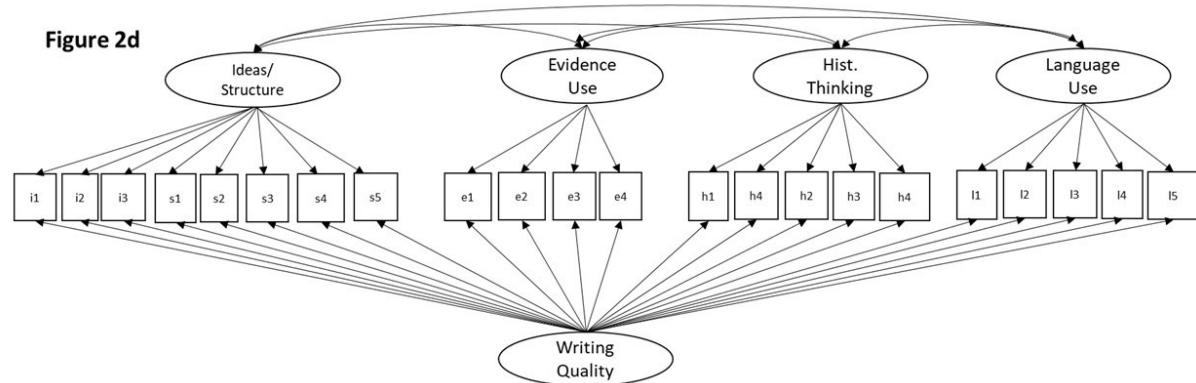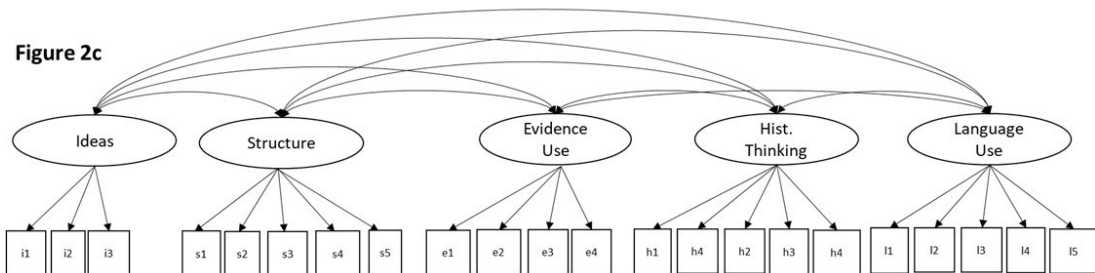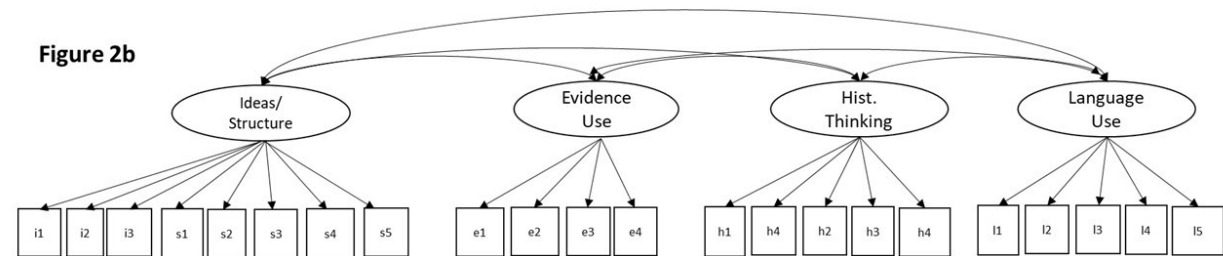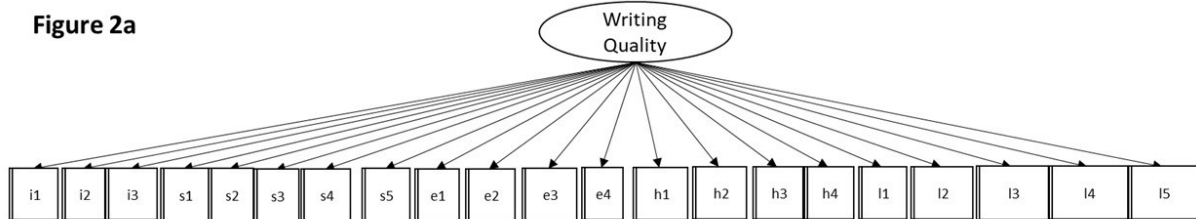
Next, the best fitting model was used in a structural regression model with each

dimension of writing predicting the holistic score. A second model included three dichotomous variables as controls—student sex, EL status, and being in high school—to assess the relative contributions of the identified dimensions to overall writing quality, controlling for these variables.

Findings

Students' writing across items in the analytic framework

My hypothesis that students' writing performance would be similar on average across discrete features of writing such as the presentation of ideas, structure, and language use was confirmed. The second part of the hypothesis that students would struggle with features of writing related to historical thinking and evidence use was partially confirmed. Students did score lower for items related to historical thinking and some items related to evidence use, but the specific item referring to use of source material for evidence showed higher performance than other items. Figure 3 shows the average performance on each analytic framework item for the 207 students in the sample

Figure 3.
Average Scores on Analytic Framework Items
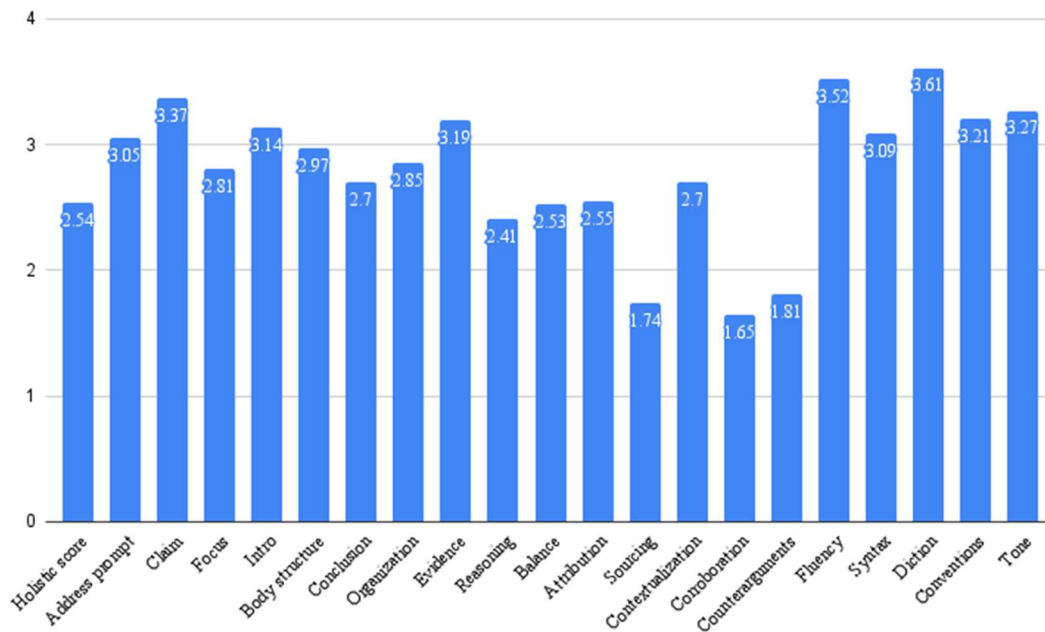
Note. All items were scored on a scale of 1-7, though the vertical axis only extends to 4 points to improve readability.

Results show that students can advance claims, integrate evidence from sources, and write an introduction more successfully than other writing skills such as reasoning with evidence, sourcing documents, and presenting and refuting counterarguments. While Figure 3 shows a broad picture of student performance across items, in what follows I describe student performance on the items across grade levels and EL status in more detail. I will then discuss the implications of relative performance across items in the general sample, by grade level, and by EL status.

Student writing differences across grade levels

There were significant grade level differences for all, but two items in the analytic framework (sourcing and contextualization), that show a somewhat linear path towards writing development for students in the sample. The grade 10 students in the sample outperformed students in higher grades on many items and students in grade 11 had lower average scores than lower grades for many items. Both these points will be discussed in the Limitations section as the

unique features of the sample could account for these findings. Table 4 shows the average score

across analytic framework items for students in grades 6, 7, and 8 respectively.

Table 4.
Descriptive Statistics of Student Writing Skills by Grade (Grades 6, 7, and 8)

	Grade 6			Grade 7			Grade 8		
	Mean	S.D.	Max	Mean	S.D.	Max	Mean	S.D.	Max
Ideas									
Address prompt	2.17	1.03	6.00	2.84	1.42	6.00	3.22	1.37	6.00
Present clear and compelling claim	2.56	1.54	6.00	3.16	1.76	6.00	3.44	1.74	7.00
Focus on substantiating claim	1.86	1.22	6.00	2.51	1.41	5.00	2.93	1.52	6.00
Structure									
Present clear introduction	2.19	1.19	6.00	2.70	1.45	6.00	3.56	1.53	6.00
Body structure	2.19	1.12	5.00	2.65	1.03	5.00	3.22	1.53	6.00
Present clear conclusion	1.89	1.30	5.00	2.30	1.56	5.00	2.59	1.80	6.00
Organization	2.03	1.08	5.00	2.49	1.22	5.00	3.04	1.43	6.00
Evidence Use									
Use source material as evidence	2.28	1.11	4.00	3.00	1.39	7.00	3.44	1.55	6.00
Commentary/reasoning	1.44	1.00	5.00	1.95	1.25	5.00	2.67	1.57	6.00
Balance purposeful summary/evidence/commentary	1.64	0.90	4.00	2.16	1.04	5.00	2.70	1.46	5.00
Attribution	1.97	1.00	5.00	2.22	0.92	5.00	2.37	1.11	5.00
Historical Thinking									
Sourcing	1.58	0.84	4.00	1.49	0.80	4.00	1.70	0.87	3.00
Contextualization	2.25	1.11	5.00	2.65	0.89	5.00	2.78	1.28	6.00
Corroboration	1.31	0.52	3.00	1.46	0.84	5.00	1.70	0.99	5.00
Address counterarguments	1.53	0.97	5.00	1.57	0.90	5.00	1.85	1.23	5.00
Language Use									
Demonstrate sentence fluency and flow	2.33	1.01	4.00	3.43	1.14	5.00	3.52	1.16	6.00
Demonstrate syntactic variety and style	2.03	0.88	4.00	3.00	0.85	4.00	3.15	1.23	6.00
Demonstrate command of diction and word choice	2.36	0.99	4.00	3.46	1.12	5.00	3.48	1.25	6.00
Demonstrate control of conventions	2.08	0.87	4.00	3.05	1.13	5.00	3.00	1.18	5.00
Use appropriate tone	2.00	1.01	4.00	3.08	1.14	5.00	3.33	1.33	7.00

Note. Scores were assigned on a scale of 1-7.

Table 5 shows grade level performance for students in grades 10, 11, and 12.

Table 5.
Descriptive Statistics of Student Writing Skills by Grade (Grades 10, 11, and 12)

	Grade 10	Grade 11	Grade 12

	Mean	S.D.	Max	Mean	S.D.	Max	Mean	S.D.	Max
Ideas									
Address prompt	3.57	1.61	7.00	2.97	1.36	6.00	3.53	1.60	7.00
Present clear and compelling claim	3.81	1.66	6.00	3.49	1.54	6.00	3.71	1.64	6.00
Focus on substantiating claim	3.35	1.72	6.00	2.86	1.46	6.00	3.26	1.68	6.00
Structure									
Present clear introduction	3.57	1.69	7.00	3.14	1.56	7.00	3.88	1.81	7.00
Body structure	3.62	1.55	7.00	2.71	1.36	6.00	3.50	1.62	6.00
Present clear conclusion	3.38	1.83	6.00	2.91	1.98	6.00	3.24	2.15	7.00
Organization	3.54	1.48	6.00	2.74	1.34	6.00	3.41	1.58	6.00
Evidence Use									
Use source material as evidence	3.73	1.68	7.00	2.89	1.32	5.00	3.68	1.61	6.00
Commentary/reasoning	2.86	1.84	7.00	2.20	1.53	6.00	3.26	1.68	6.00
Balance purposeful summary/evidence/commentary	2.92	1.77	7.00	2.43	1.52	7.00	3.26	1.71	6.00
Attribution	3.35	1.78	7.00	2.34	1.24	6.00	3.06	1.46	6.00
Historical Thinking									
Sourcing	2.05	1.08	5.00	1.80	1.16	5.00	1.82	0.90	4.00
Contextualization	3.00	1.27	5.00	2.63	1.19	5.00	2.91	1.08	6.00
Corroboration	1.97	0.96	4.00	1.54	0.82	4.00	1.91	1.03	4.00
Address counterarguments	2.43	1.39	6.00	1.77	0.97	5.00	1.74	1.05	5.00
Language Use									
Demonstrate sentence fluency and flow	4.19	1.41	7.00	3.37	1.44	7.00	4.03	1.71	7.00
Demonstrate syntactic variety and style	3.68	1.43	7.00	3.06	1.41	7.00	3.44	1.40	6.00
Demonstrate command of diction and word choice	4.24	1.46	7.00	3.51	1.56	7.00	4.35	1.84	7.00
Demonstrate control of conventions	3.89	1.41	7.00	3.14	1.52	7.00	3.82	1.71	7.00
Use appropriate tone	4.03	1.59	7.00	3.09	1.62	7.00	3.88	1.92	7.00

Note. Scores were assigned on a scale of 1-7.

My original hypothesis was that differences across language use items would be relatively muted compared to differences in other items, especially for older students. However, this is not reflected in the data which show variation across many items and grade levels, with the most consistent grade level differences occurring between grade 6 students and other students in the sample. To make the above results more easily comprehensible, I will describe grade level differences by categories enumerated in the hypothesized dimensions.

Presentation of Ideas

Figure 4 shows grade level performance for items related to the presentation of ideas.

Figure 4.
Grade Level Performance for Items Related to the Presentation of Ideas
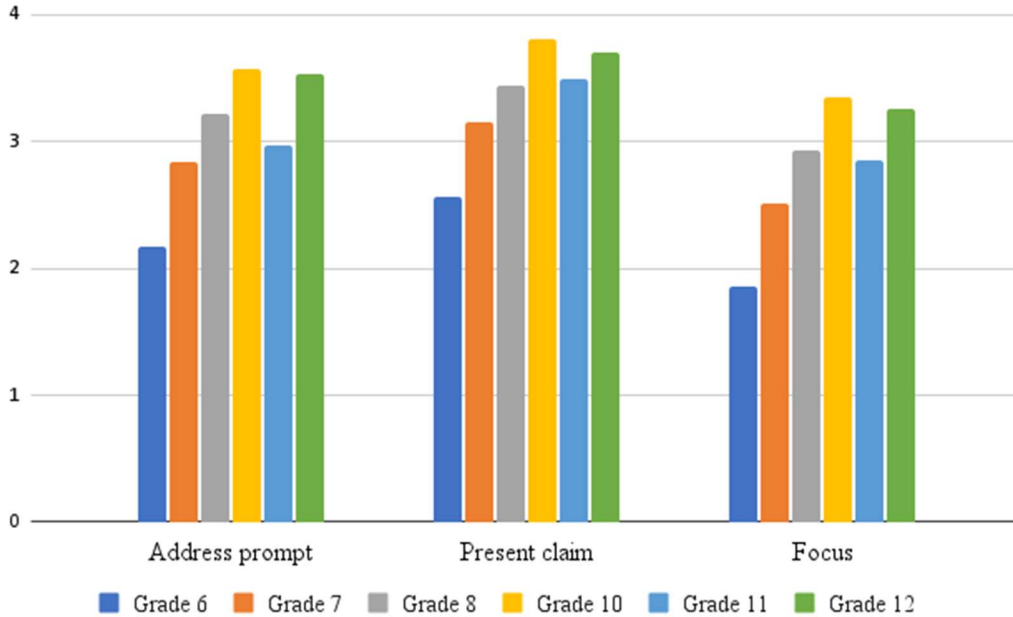
Note. Scores were assigned on a scale of 1-7.

For each analytic framework item, a one-way ANOVA was performed to see the effect of grade level on that item, followed by a post-hoc test using a Bonferroni adjustment. One-way ANOVA revealed significant differences in mean scores for at least two groups for the item "Addressing aspects of the prompt" ($F_{(5, 202)} = 4.87$, $p < 0.01$). Post-hoc tests revealed students in grade 6 scored lower than all other grade levels and students in grade 7 performed significantly worse than students in grade 10 and 12. For the items "Presenting a clear and compelling claim" and "Focus on substantiating claim" there was only a statistically significant difference for students in grade 6 and students in grade 8, 10, 11, and 12 ($F_{(5, 202)} = 2.736$, $p < 0.01$ and $F_{(5, 202)} = 4.738$, $p < 0.01$).

Overall, students' abilities to address aspects of the prompt, advance a clear claim, and focus on proving that claim grew substantially from grade 6 to 8. Comparing grade 6 and 8

students for addressing the prompt, for example, we see a mean score difference of 1.05. The growth for students in grades 10, 11, and 12 is not evident, however, for these items.

Structure

Figure 5.
Grade Level Performance for Items Related to Structure
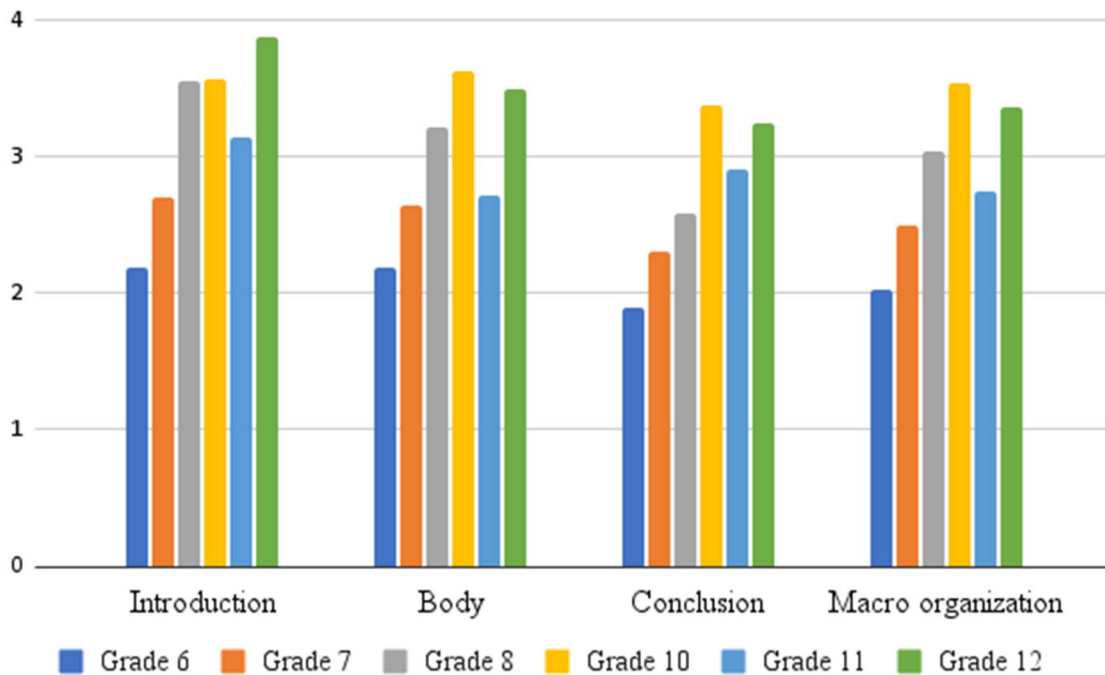
Using ANOVA, we find statistically significant grade level differences for "Presenting a clear introduction" ($F_{(5, 202)} = 5.797$, $p < 0.01$), with students in grades 6 and 7 scoring lower than grades 8, 10, and 12. The item measuring the quality of the conclusion also indicated notable grade level differences ($F_{(5, 202)} = 3.659$, $p < 0.05$), with grades 6 and 7 students scoring lower than grades 10, 11, and 12 students.

For the item measuring the structure of the body, there were many differences between grade levels ($F_{(5, 202)} = 5.776$, $p < 0.01$): Grade 6 students scored lower than grades 8, 10, and 12 students (but not grade 11 students); grade 7 students scored lower than grade 10 and 12 students, and grade 10 students score lower than grade 11 students. The item measuring macro

organization showed significant differences (F(5, 202) = 6.380, p < 0.01) with grade 6 students

scoring lower than all other students except grade 7, grade 7 students scoring lower than grades

10 and 12, and grade 11 students scoring lower than grades 10 and 12.

Overall, students' varied widely across skills related to structuring an argumentative

essay, with significant growth across grades 6-8 that sometimes extended to grades 10

(introduction) and 12 (body and macro organization). The differences between grade levels were

also substantial. Grade 12 students scored on average 1.6 points higher in presenting an

introduction, which is the largest difference for any item in the analytic framework. In sum, all

items related to structure appear quite malleable throughout development, though muted

differences occurred between secondary grades for some items.

Evidence Use

Next, I examined grade level differences for items related to Evidence Use. Figure 6

shows the average score for analytic items by grade level for students in the sample.

Figure 6.
Grade Level Performance for Items Related to the Evidence Use
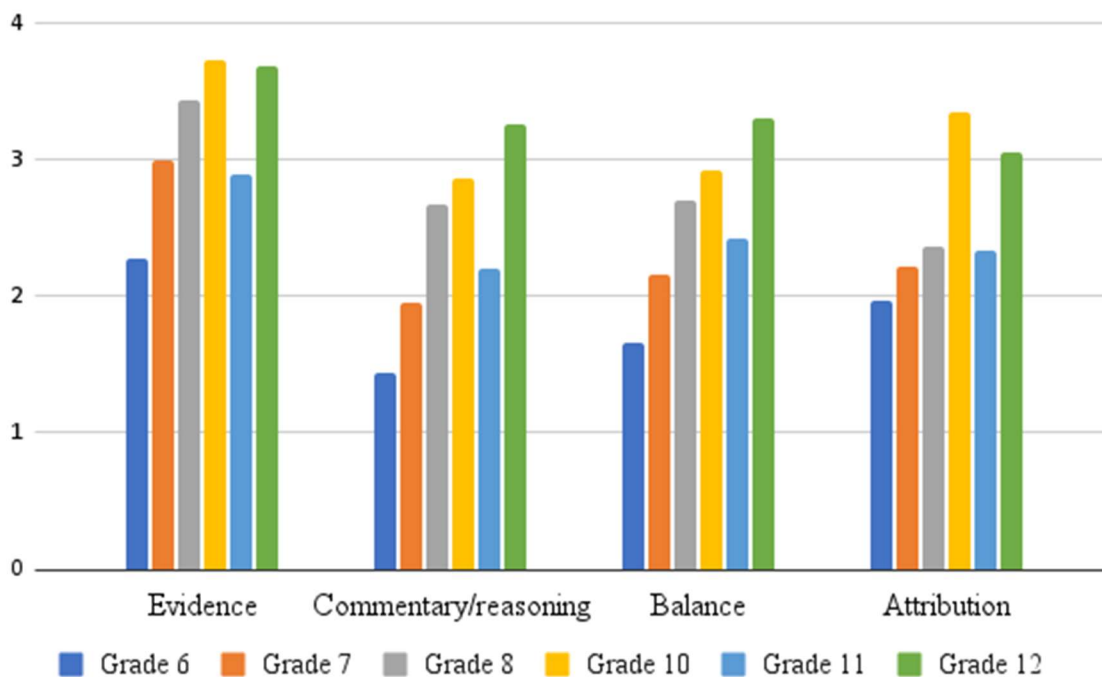
Using ANOVA, we find statistically significant grade level differences for "Use source material as evidence: ($F(5, 202) = 5.198$, $p < 0.01$), with students in grade 6 scoring lower on average than all grades except grade 11; students in grade 7 scoring significantly lower than students in grade 10; and grade 11 students scoring significantly lower than grade 10 and 12 students.

For the items measuring the quality of commentary/reasoning and the balance of purposeful summary, commentary and reasoning, there were equivalent patterns of statistically significant level differences ($F(5, 202) = 6.868$, $p < 0.01$ and $F(5, 202) = 5.703$, $p < 0.01$ respectively). For both items, grade 6 students had lower scores than grade 8, 10, 11, and 12 students; grade 7 students has lower scores than grade 10 and 12; and grade 11 students scored lower on average than grade 12 students.

A final item measured the quality of attribution. There was less overall variation across grades for this item, namely students in grades 10 and 12 performed higher on average than other

50

grade levels ($F(5, 202) = 6.120$, $p < 0.01$). Many students did not successfully reference which source selected text or evidence came from; instead, they frequently mention "the text." There is evidence some grade 10 and 12 students make more clear references, for example, "source 2," but the mean scores indicated a lack of specific references to where evidence came from (e.g. "Chavez's letter to the public"), which is an important part of argument writing in history.

For the most part, patterns in grade level differences for Evidence Use resembled patterns for Ideas and Structure with large growth in scores across middle school students. For example, grade 10 students scored on average 1.16 points higher than grade 6 students in using source material. For one item, the use of source material, there were less pronounced differences across items with high school students. Though grade 12 students did not significantly outperform lower grades in other items in the Structure and Ideas categories, they did so for commentary/reasoning and the balance of purposeful summary, evidence, and commentary. For example, the difference in the quality of commentary/reasoning between grades 6 and 12 was 1.82 points on average.

Historical Thinking

Grade level differences in writing components to historical thinking are described next. Figure 7 shows the average performance in items across grade levels.

Figure 7.
Grade Level Performance for Items Related to Historical Thinking
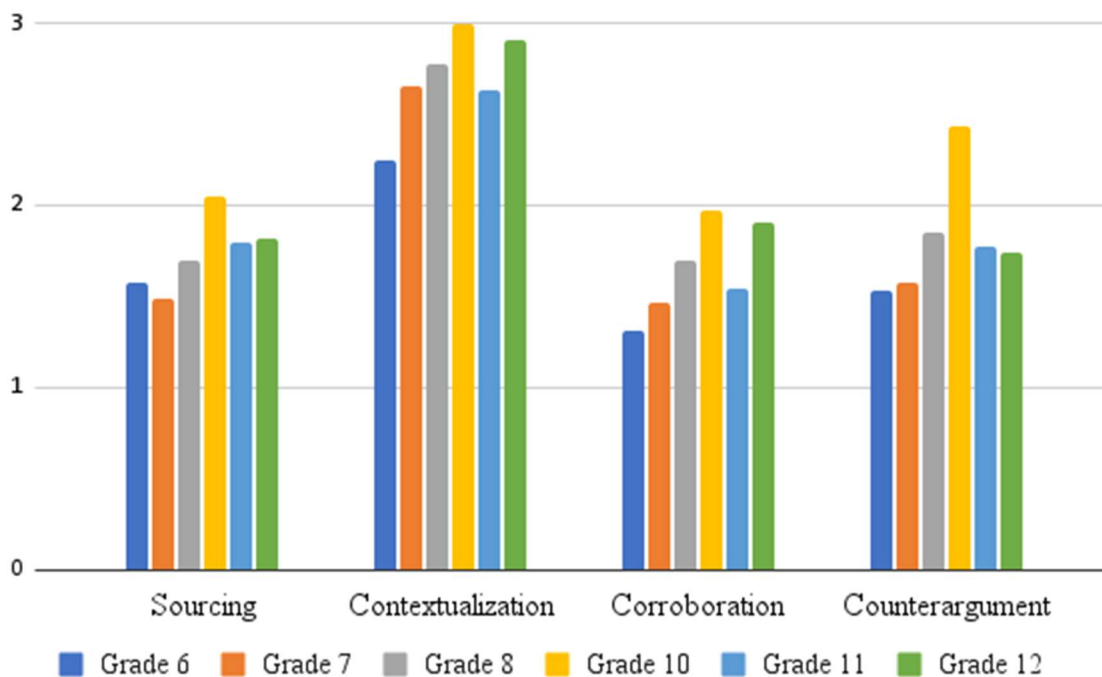
Using ANOVA, there were no statistically significant grade level differences for sourcing ($F(5, 202) = 1.606$, $p = 0.16$) and contextualization ($F(5, 202) = 1.945$, $p = 0.09$). The average scores for all students in the sample differed across these items, with the average score for sourcing (1.74) being close to a full point lower than the average score for contextualization (2.70). Reasons for this difference are discussed in the following section summarizing major differences across grade levels.

ANOVA did find statistically significant differences for corroboration ($F(5, 202) = 3.256$, $p < 0.01$), with grade 6 and 7 students having lower scores than grade 10 and 12 students and grade 11 students having lower scores than grade 10 students. Differences also existed for "Addressing and refuting a counterargument" ($F(5, 202) = 3.286$, $p < 0.01$) with grade 10 students scoring higher than all other grades.

The outperformance of grade 10 students is notable. Classroom level differences, not necessarily grade level, likely contributed to this variation, especially with only three teacher's

students per grade featured in this study. Though there are some apparent, non-statistically significant differences across grades, many historical thinking items feature less variation than other analytic items.

Language use

Lastly, we consider grade level differences for items related to language use, as illustrated by Figure 8.

Figure 8.
Grade Level Performance for Items Related to Language Use
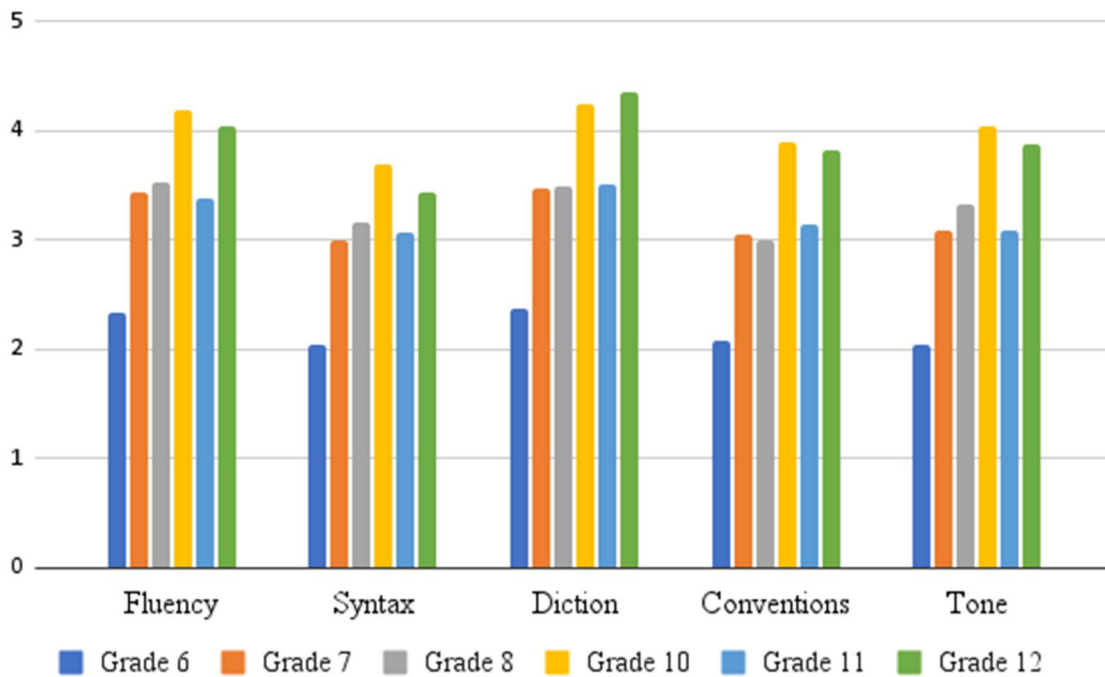
Using ANOVA, we can see statistically significant grade level differences for "Sentence fluency and flow" ($F(5, 202) = 8.607$, $p < 0.01$), with grade 6 students performing lower than all other grades, a pattern repeated for all other items within this category. Additionally, grades 7, 8, and 11 students had lower scores than grade 10, and grade 11 students had lower scores than grade 12 students. For "Syntactic variety and style ($F(5, 202) = 7.720$, $p < 0.01$), there was a similar lower performance in grade 6 students. Other differences include the lower performance

of grades 7 and 11 students compared to grade 10 students.

For the quality of diction or word choice, significant differences ($F(5, 202) = 9.296$, $p < 0.01$) include grade 6 students' lower scores, grades 7 and 8 students scoring lower than grades 10 and 12, and grade 11 students scoring lower than grade 12 students. For the item "Demonstrate control of conventions" ($F(5, 202) = 8.735$, $p < 0.01$) grade 6 students again had lower scores than all other grades, and grades 7, 8, and 11 had lower scores than grades 10 and 12. Finally, for the item measuring the use of appropriate tone ($F(5, 202) = 8.770$, $p < 0.01$), there were differences between grade 6 students and all other grades, and grade 7 and 11 students scored lower than grades 10 and 12 students.

For items related to students' language use, there is a similar pattern in growth from grades 6 to 8 and from grade 8 to 12 with grade 10 students scoring lower than would otherwise be expected if a linear progression of skills across grade levels occurred. Further, there is an interesting similarity with the language use of grade 7 and grade 8 students, with further growth seen between grades 8 and 10. A lot of language development occurs in grade 6, but less was apparent in grade 7 for the present population of students.

Grade level differences summarized

Across all items, the major differences in student writing appear across students in the middle school grades with less variation among analytic items for high school students, albeit with some exceptions (e.g. commentary/reasoning, balance, and presenting an introduction). Grade 6 students are largely in a knowledge-telling period (Scardamalia, & Bereiter, 1987); their lower scores for commentary/reasoning and focus on substantiating a claim suggest most students are summarizing or restating what they read from source texts. Grade 7 students appear to continue this growth in moving towards more complex reasoning, as well as learning the

structure of an academic essay and how to use evidence to support claims.

Some notable issues warrant further investigation with a larger sample of students from more classrooms. Sources of the lower scores for grade 11 students compared to other high school students could be unique classroom characteristics for the grade level. Another finding, namely, the lack of variation in historical thinking, warrants further investigation as well. Different writing prompts may elicit different types and occurrences of historical reasoning (De La Paz & Monte-Sano, 2012); so a claim about students' lacking these skills should be robustly tested using different writing prompts.

Presently, I hypothesize the full point difference in sourcing and contextualization scores is due to both skill differences and prompt effects. Sourcing is similar to attribution, noting where evidence comes from, but goes further to explain why the source of information matters for analysis. The low scores for attribution suggest students mostly do not know when or how to use these skills in historical argument writing. At the same time, the prompt itself may require less consideration of the sources of evidence and attention to historical context to make an argument about causality. Because casualty involves putting events in order or considering why some actor or event mattered given the surrounding circumstances, contextualization is necessary to answer the question.

Sourcing is also necessary because historical reasoning means considering where evidence comes from when making arguments. However, the need to source may be less obvious to students who consistently receive reliable sources or textbook excerpts from their teachers. Further, sourcing is often generated by a disagreement across sources, which is not salient in the prompts used presently (Bråten et al., 2017; Braasch & Bråten, 2017). Other prompts that the research team tested in earlier phases of development, "Who is responsible for the violence that

occurred at the 'Boston' Massacre'?" elicited more sophisticated sourcing moves from the students because there were two explicitly conflicting accounts of events advanced by two competing groups (i.e. the Colonists vs. the British soldiers). Therefore, prompt differences might also contribute to these score differences.

Student writing differences across EL Status

My hypothesis that students would perform lower across most analytic framework items was confirmed. There were significant differences for a broad range of items, many related to language use (e.g. sentence fluency and diction). Differences between students designated as EL and other students are summarized in Table 6 and described in greater detail in the following section.

Table 6.
Student Writing Differences by EL Status

	EO & IFEP students			RFEP students			EL students		
	Mean	S.D.	Max	Mean	S.D.	Max	Mean	S.D.	Max
Ideas									
Address prompt	3.14	1.51	7	3.20	1.47	6	2.50	1.29	5
Present clear and compelling claim	3.34	1.67	7	3.84	1.58	6	2.65	1.67	5
Focus on substantiating claim	2.89	1.63	6	3.04	1.56	6	2.09	1.22	4
Structure									
Present clear introduction	3.21	1.68	7	3.48	1.56	7	2.41	1.4	5
Body structure	3.03	1.43	6	3.21	1.64	7	2.38	1.02	4
Present clear conclusion	2.83	1.92	7	2.88	1.83	6	2.18	1.59	5
Organization	2.95	1.48	6	3.09	1.46	6	2.24	1.1	4
Evidence Use									
Use source material as evidence	3.21	1.55	7	3.50	1.56	7	2.50	1.24	5
Commentary/reasoning	2.46	1.67	7	2.70	1.66	6	1.65	1.01	5
Balance purposeful summary/evidence/commentary	2.56	1.51	7	2.88	1.68	7	1.82	1.06	5
Attribution	2.64	1.39	7	2.82	1.44	7	1.88	0.91	6
Historical Thinking									
Sourcing	1.75	0.95	5	2	1.08	5	1.29	0.58	3
Contextualization	2.77	1.16	6	2.84	1.17	6	2.24	0.96	5
Corroboration	1.66	0.91	5	1.89	0.97	4	1.24	0.5	3

Address counterarguments	1.85	1.15	5	1.93	1.23	6	1.47	0.71	3
Language Use									
Demonstrate sentence fluency and flow	3.56	1.48	7	3.57	1.43	6	3.06	1.37	5
Demonstrate syntactic variety and style	3.20	1.37	7	3.05	1.24	5	2.59	1.16	5
Demonstrate command of diction and word choice	3.66	1.54	7	3.68	1.6	7	3.12	1.37	5
Demonstrate control of conventions	3.29	1.49	7	3.30	1.43	6	2.56	1.24	6
Use appropriate tone	3.32	1.62	7	3.41	1.65	7	2.62	1.3	6

Note. 117 students were designated as EO and IFEP; 56 students were designated as RFEP; 34 students were designated as EL.

Similar to the previous section, I will describe differences between EL designations by the categories enumerated in the hypothesized dimensions for SBAW in history.

Presentation of Ideas

Figure 9 shows differences between students designated as EO/IFEP, RFEP, and EL for items related to the presentation and development of ideas.

Figure 9.
Performance across EL Status for Items Related to Presentation of Ideas
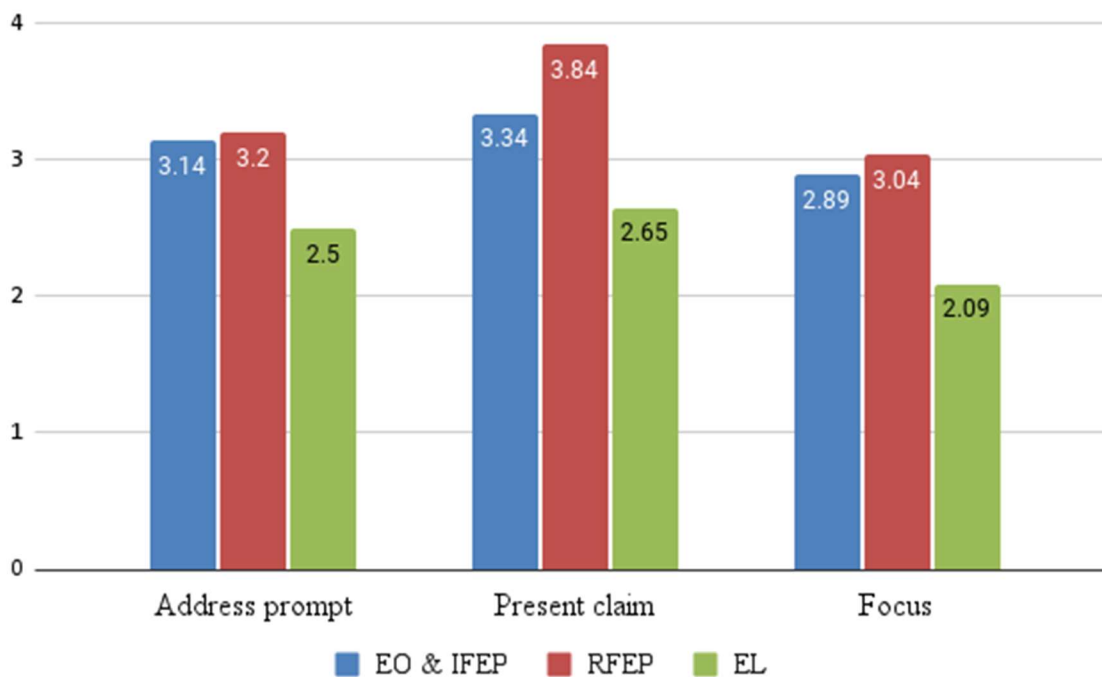
For each analytic framework item, a one-way ANOVA was used to compare the effect of EL Status on each item, with post-hoc tests using a Bonferroni adjustment showing for which grade levels there were significant mean differences. For the item, "Addresses all aspects of the prompt," it appears EL students score lower. However, using ANOVA, differences between EL students and other students were insignificant: ($F_{(2, 205)} = 2.87$, $p = .06$), most likely due to sample size limitations. There were statistically significant differences by EL status for the items "Presenting a clear and compelling claim" ($F_{(2, 205)} = 5.58$, $p < .01$) and "Focus on substantiating claim" ($F_{(2, 205)} = 4.41$, $p < .05$). While RFEP students had an average score of 3.84 for presenting a claim, EL students had an average score of 2.65—1.19 points lower. The significant difference for "Focus on substantiating a claim" was also between students designated RFEP and EL, with EL students scoring 0.95 points lower on average. These findings suggest many students designated as English learners are challenged by presenting their ideas in writing

as would be expected, though the performance of their RFEP peers suggests they are capable of growing in these skills.

Structure

Figure 10 shows performance on items related to the structure or organization of an essay.

Figure 10.
Performance across EL Status for Items Related to Structure
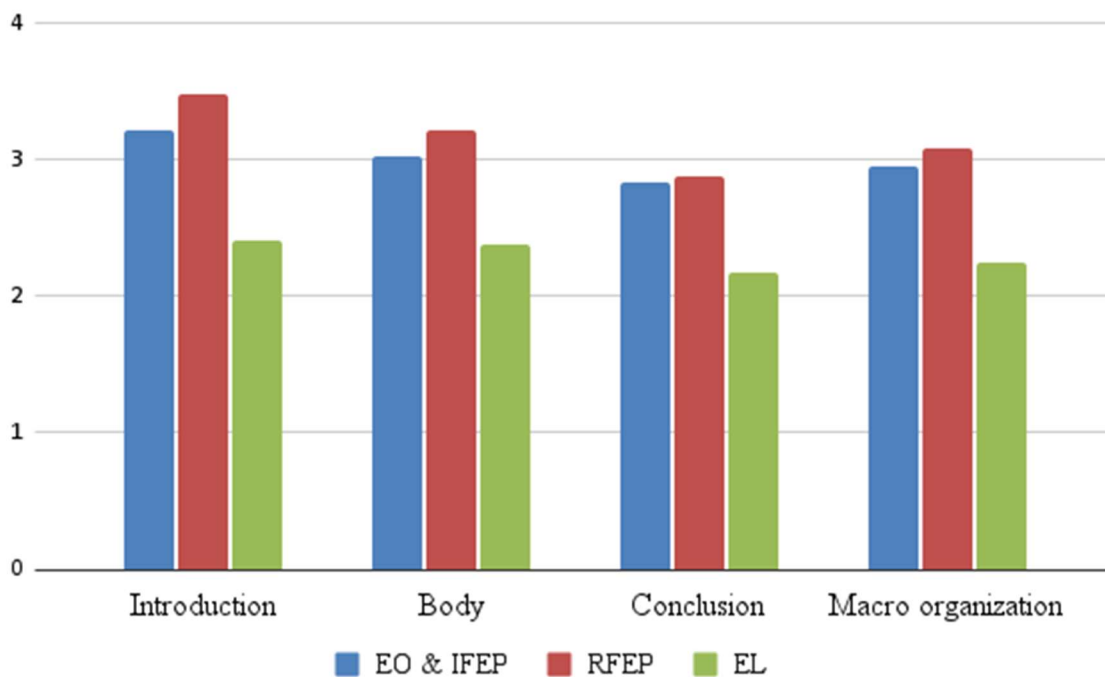
Using ANOVA, we can see statistically significant differences across EL status for "Presenting a clear introduction" ($F2, 205$) = 4.90, $p < 0.01$), with EL students scoring significantly lower for both REFP and EO/IFEP students. The item measuring the structure of the body also saw notable differences ($F(2, 205) = 3.78$, $p < 0.05$), with EL students performing 0.83 points lower on average.

 For the item measuring the conclusion, there was no statistically difference across EL status ($F(2, 205) = 1.87$, $p = .15$). The item measuring organization, how well the overall essay

was organized including transitions between major sections, showed significant differences (F(2, 205) = 4.23, p < 0.05) with EL students scoring significantly lower than both REFP and EO/IFEP students.

Overall, students' varied significantly across skills related to structuring an argumentative essay, though all students were challenged by writing a conclusion. It is important to note that students only had 50 minutes to write their argumentative essays. Additionally, it is important to note there were no significant differences between RFEP and EO/IFEP students for these items, a similar pattern to those items related to the presentation of ideas. Though insignificant statistically, there also appears to be some outperformance by RFEP students.

Evidence Use

Next, we examine differences for items related to Evidence Use. Figure 11 shows the average score for analytic items by EL status for students in the sample.

Figure 11.
Performance across EL Status for Items Related to the Evidence Use
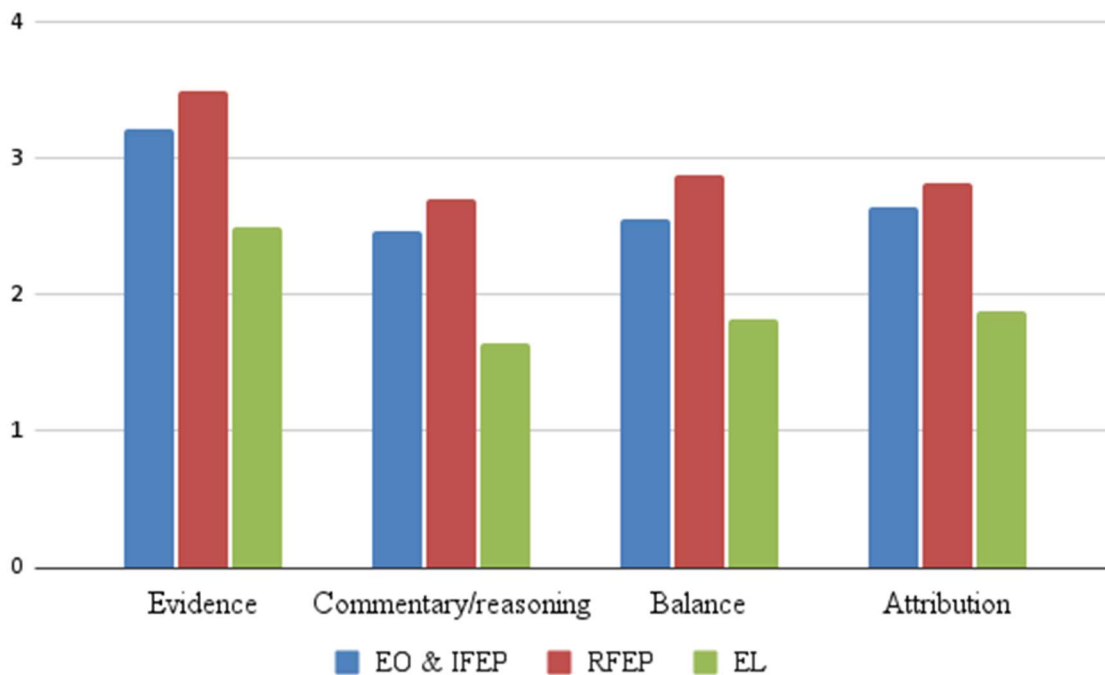
Using ANOVA, we can see statistically significant differences for "Use source material as evidence:" ($F(2, 205) = 4.73$, $p < 0.01$), with EL students scoring lower on average than RFEP students. For the items measuring the quality of commentary/reasoning and the balance of purposeful summary, commentary and reasoning, there were equivalent patterns of statistically significant differences ($F(2, 205) = 4.94$, $p < 0.01$ and $F(1,2\ 07) = 528$, $p < 0.01$ respectively). For both items, EL students scored lower than their RFEP and EO/IFEP peers. The same pattern held for the attribution of evidence ($F(2, 205) = 5.62$, $p < 0.01$). Overall, differences across EL status for items related to evidence use resembled differences for other items. Integrating evidence, providing reasoning and commentary, and attributing evidence to sources are particularly challenging for English learners.

Historical Thinking

Differences in writing components related to historical thinking are described next. Figure 12 shows the average performance in items across EL status.

Figure 12.
Performance across EL Status for Items Related to Historical Thinking
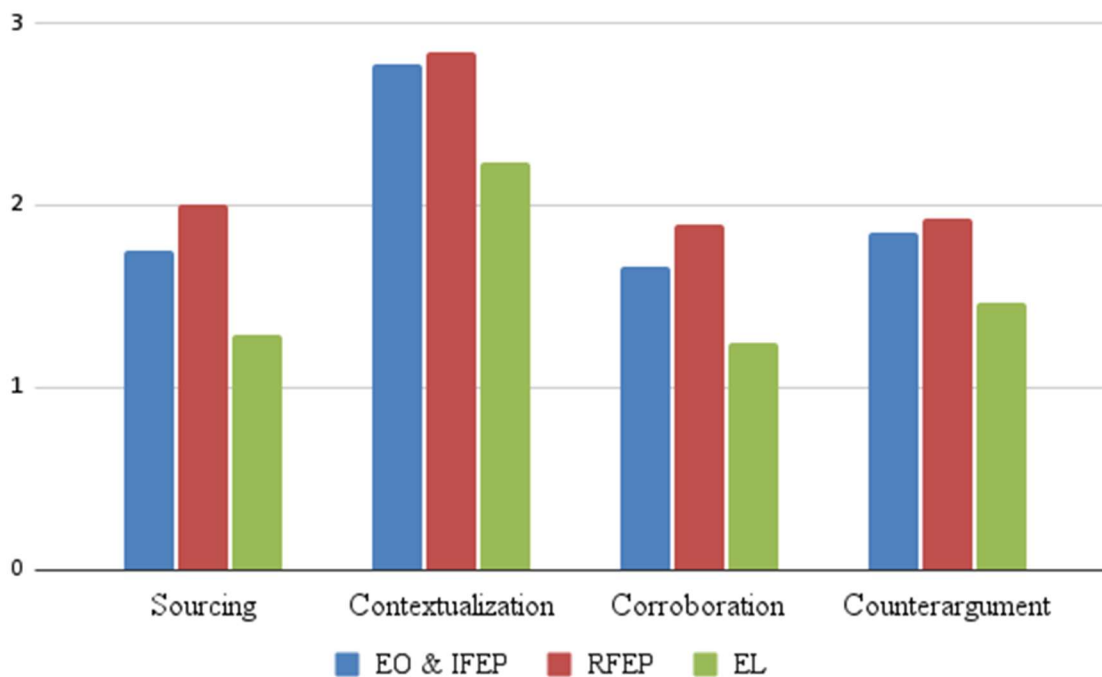
Using ANOVA, there were no statistically significant grade level differences for "Presenting and addressing a counterargument" ($F_{(2, 205)} = 1.98$, $p = 0.14$). For sourcing ($F_{(2, 205)} = 6.02$, $p < .01$), contextualization ($F_{(2, 205)} = 3.49$, $p < .05$), and corroboration ($F_{(2, 205)} = 5.99$, $p < .01$) there were significant differences with students designated as EL scoring lower on average than both their RFEP and EO/IFEP peers. In sum, while all students were challenged to integrate historical thinking moves in their writing, especially sourcing, EL students were particularly challenged.

Language use

Lastly, we consider differences for items related to language use across EL status. See Figure 13.

Figure 13.
Performance across EL Status for Items Related to the Language Use
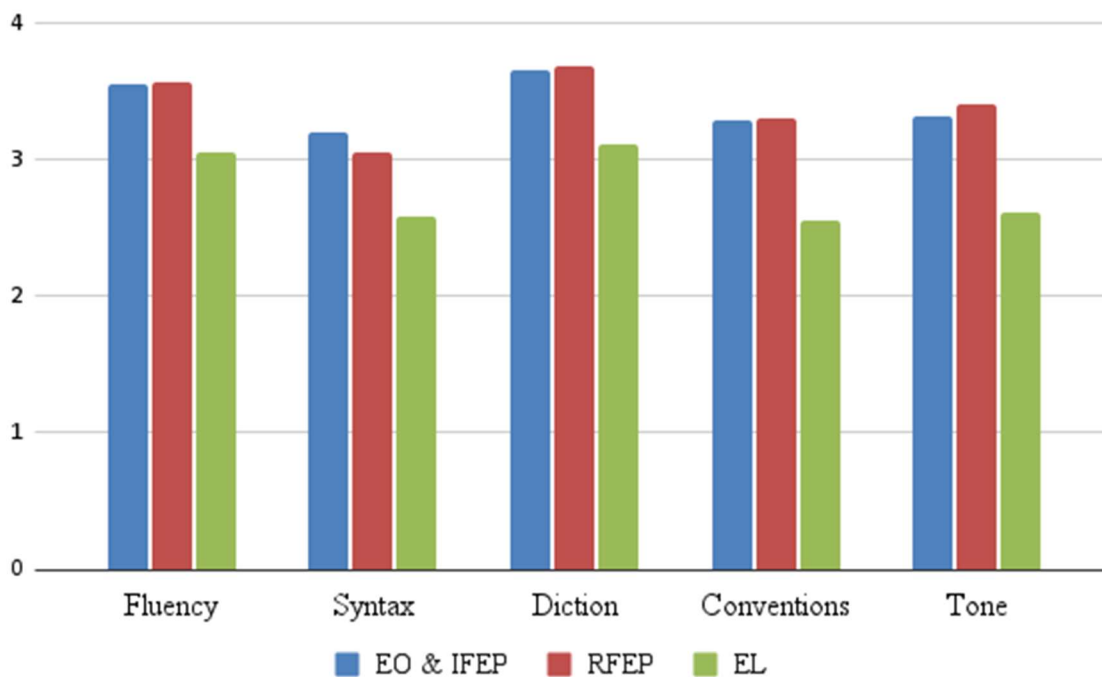
First, there were no differences between RFEP students in their EO/IFEP peers. Further, using ANOVA, there were no statistically significant differences across EL status for "Sentence fluency and flow" ($F_{(2, 205)} = 1.74$, $p = .17$), "Syntactic variety and style" ($F_{(2, 205)} = 2.86$, $p = .06$), and "Demonstrate command of diction and word choice" ($F_{(2, 205)} = 1.83$, $p = .16$). Differences were found for "Demonstrate control of conventions" ($F_{(2, 205)} = 3.75$, $p < 0.05$) and "Uses appropriate tone" ($F_{(2, 205)} = 3.13$, $p < 0.05$), with EL students scoring lower on average than their RFEP peers. The muted differences between students classified as EL for some items are notable given the magnitude of differences for other items and the nature of these items that largely assess proficiency with academic language.

Differences across EL Status Summarized

Across all items, major differences appeared in writing across students designated as EL and students designated as RFEP with some exceptions (e.g. fluency, diction, syntax). Students designated as EL experienced challenges in many aspects of writing, especially the presentation

of ideas and using evidence. The performance of RFEP students as well as extant research suggests these students can grow in these areas when instruction focuses on helping students engage in knowledge transformation and higher-order thinking around texts (Olson et al., 2017; Olson et al., 2020). However, current challenges in using the English language to communicate can serve as a barrier to learning or applying higher-level thinking skills in writing.

The low sample size (117 students were designated as EO and IFEP; 56 students were designated as RFEP; 34 students were designated as EL) is worth noting. Some non-significant differences might become significant in a larger sample. Additionally, examining differences in EL status by grade level might offer a more detailed picture of EL students' development in their language skills over time. Such analyses were not feasible given the size of the current sample.

Dimensions of source-based arugment writing in history

To better understand the features of source-based arugment writing in history, data from the analytic framework and CFA, using Mplus 8.4 (Muthén & Muthén, 2017), were used to examine the dimensions of SBAW in history.

The normality and variation of variables in the analytic framework were assessed before conducting CFA. The distribution of scores, skewness, and kurtosis were all adequate with the exceptions of attribution, sourcing, corroboration, and counterarguments. Table 7 shows the properties of variables with an asterisk indicating variables with unsatisfactory distributions and normality.

Table 7.
Descriptive Statistics for Items in Analytic Framework and Holistic Score

	Mean	SD	Minimum	Maximum	Skewness	Kurtosis
Holistic	2.57	1.20	1.00	6.00	0.51	2.51
Address prompt	3.05	1.48	1.00	7.00	0.59	2.29
Claim	3.36	1.68	1.00	7.00	-0.25	1.70
Focus	2.80	1.58	1.00	6.00	0.44	2.06

Introduction	3.15	1.63	1.00	7.00	0.54	2.28
Body	2.98	1.45	1.00	7.00	0.37	2.39
Conclusion	2.73	1.85	1.00	7.00	0.50	1.71
Organization	2.87	1.44	1.00	6.00	0.38	2.27
Evidence	3.17	1.53	1.00	7.00	0.19	2.21
Commentary	2.39	1.61	1.00	7.00	0.87	2.48
Balance	2.52	1.53	1.00	7.00	0.92	2.89
Attribution	2.57	1.37	1.00	7.00	1.07	3.58
Sourcing*	1.74	0.96	1.00	5.00	1.19	3.81
Contextualization*	2.70	1.15	1.00	6.00	0.39	2.80
Corroboration*	1.65	0.89	1.00	5.00	1.43	4.66
Counterargument*	1.81	1.12	1.00	6.00	1.43	4.37
Fluency	3.48	1.45	1.00	7.00	0.01	2.47
Syntax	3.06	1.32	1.00	7.00	0.31	2.87
Diction	3.57	1.53	1.00	7.00	0.06	2.44
Conventions	3.17	1.45	1.00	7.00	0.35	2.58
Tone	3.23	1.60	1.00	7.00	0.49	2.71

Note. All items were scored on a scale of 1 to 7.

The right-skew of the variables indicated with an asterisk was likely due to exceptionally, low, but accurate scores as these types of thinking are particularly difficult for students to master, as discussed previously. Therefore, logarithmic transformations were conducted since excessive skew can violate the assumptions of uni- and multivariate normality needed for maximum likelihood functions in CFA (Kline, 2015). After transformation, the variables had acceptable skew to be used in CFA.

Table 8.
Descriptive Statistics for Items with Logarithmic Transformations

	Mean	SD	Minimum	Maximum	Skewness	Kurtosis
Attribution	0.81	0.52	0.00	1.95	0.05	2.33
Sourcing	0.43	0.49	0.00	1.61	0.58	1.91
Contextualization	0.38	0.47	0.00	1.61	0.73	2.23
Counterargument	0.44	0.53	0.00	1.79	0.73	2.19

Bivariate correlations between variables are also presented in Table 9.

Table 9.
Bivariate Correlations for Items in Analytic Framework

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1. address	--																			
2. pclaim	.73	--																		
3. focus	.87	.82	--																	
4. intro	.57	.59	.60	--																
5. body	.77	.66	.84	.52	--															
6. conc	.72	.58	.76	.45	.65	--														
7. org	.81	.70	.88	.68	.89	.82	--													
8. evidence	.77	.68	.82	.51	.83	.69	.84	--												
9. commentary	.76	.59	.80	.50	.80	.62	.78	.75	--											
10. balance	.74	.60	.79	.50	.81	.64	.78	.82	.93	--										
11. attrib	.57	.44	.60	.46	.72	.49	.67	.64	.61	.63	--									
12. sourcing	.52	.43	.57	.37	.63	.49	.62	.57	.57	.56	.67	--								
13. context	.65	.58	.69	.39	.72	.63	.72	.72	.66	.66	.53	.64	--							
14. corrob	.60	.49	.62	.40	.65	.56	.66	.64	.62	.61	.61	.72	.72	--						
15. counter	.56	.44	.59	.34	.60	.57	.62	.57	.54	.53	.38	.55	.72	.68	--					
16. fluency	.71	.57	.72	.53	.78	.63	.79	.75	.64	.68	.60	.48	.64	.55	.49	--				
17. syntax	.72	.58	.73	.48	.77	.68	.78	.74	.66	.68	.60	.49	.68	.57	.52	.88	--			
18. diction	.67	.53	.68	.45	.74	.63	.73	.71	.64	.68	.58	.48	.67	.56	.49	.90	.89	--		
19. convention	.65	.54	.66	.46	.72	.59	.72	.68	.59	.63	.61	.47	.60	.57	.48	.89	.87	.89	--	
20. tone	.69	.53	.69	.49	.74	.59	.73	.70	.65	.68	.65	.53	.60	.59	.47	.86	.84	.85	.88	--

Note. All coefficients were statistically significant at the .001 level.

Examining the bivariate correlations, all variables were moderately or strongly related with each other. After further examination, I decided the variable "org," which measures the macro organization of the writing, should be removed from analyses. Empirically, this item was strongly correlated with many variables and closely resembled the bivariate correlations of another variable "body," which measured how well the body of the essay was organized. Therefore, even if "org" was dropped from analyses, much of the variance explained by the variable was already represented in "body." Substantively, the coders of this variable also reported that the variable prompted a very "holistic" evaluation of writing—as if it were a sum of all other variables and, therefore, not appropriate to the purpose of the analytic framework. Initial CFAs were conducted, which confirmed the variable was causing issues of multicollinearity.

Next, to determine the dimensions of SBAW in history, the four competing models in

Figure 2 were fitted to the data with the model fits reported in Table 10.

Table 10.
Model Fits of Confirmatory Factor Analysis Shown in Figure 2

	χ^2, DF; p-value	CFI	RMSEA	SRMR	Model comparison
Figure 2a	1286.97; 153; <.001	.757	.189	.106	--
Figure 2b	473.59; 146; <.001	.931	.104	.042	2b vs. 2a
					$\Delta\chi^2 = 813.38, \Delta df = 7, p < .001$
Figure 2c	**442.446; 142; <.001**	**.936**	**.101**	**.051**	2c vs. 2b
					$\Delta\chi^2 = 31.145, \Delta df = 4, p < .001$
Figure 2d*	--	--	--	--	--

Note. *For the bifactor model, there was no convergence. The model in bold was the best-fitting model, according to fit statistics but was not selected as the model for further analyses.

First, there were convergence issues with the bifactor model. Attempts to resolve non-convergence included removing indicators and fixing parameters such as correlations to set values, but convergence issues were not resolved.

Next, chi-square difference test indicated that the correlated four-factor model (Figure 2b) had good fit and superior fit to the unidimensional model (Figure 2a) ($\Delta\chi^2 = 813.38, \Delta df = 7$, $p < .001$). The chi-square difference test also indicated that correlated five-factor model (Figure 2c) had superior fit to the four-factor ($\Delta\chi^2 = 31.145, \Delta df = 4, p < .001$.) Based on model fit indices, the two models had good fit, with the chi-square test suggesting the five-factor model should be preferred.

The four-factor model was selected for further analyses and respecification due to the following reasons. First, previous research examining the dimensions of writing in this area suggested that Ideas and Structure might be a single factor (Steiss, et al., 2022); The ideas of an argument essay need to be structured well to be communicated effectively; therefore, these two might not be dissociable. This claim is also supported by the empirical analyses in the present study for two reasons: 1) the Ideas and Structure factors in the five-factor model were correlated at 0.996, which is so high as to suggest that these are *not* distinct and dissociable constructs in

writing and 2) multiple modification indices suggested substantial improvement in model fit if cross loadings were permitted by several indicators from the Ideas and Structure factors. Finally, the four-factor model still had good fit (Hu and Bentler, 1999; Kline 2015)

At this point, an alternative model was fit to the data based on the fit of previous models, item correlations and loadings, and understanding of the analytic framework. First, the item "body" was set to load onto both the Evidence Use and Ideas/Structure factors. Bivariate correlations as well as discussions with coders using the analytic framework suggested both the latent factors of Structure and Evidence use were possibly being measured with this item. While structuring the body of an essay can be plausibly due to a writer's organization skills, it also requires skills in using evidence and commentary as these are the substantive components of the body. Therefore, the item "body" was set to load onto both factors.

Second, correlated residuals were allowed between four item pairs which reflect overlap in how these items were measured and thus relate for reasons other than the shared factor—balance with commentary, attribution with sourcing, intro with pclaim and focus with pclaim. Within all these pairs, measuring one necessarily means measuring part of another. For example, some attribution to the source material is needed for a student to engage in sourcing. From this point, students can "source" documents with varying degrees of proficiency. Similarly, the quality of an introduction is directly dependent on the presence of a clear claim even though these are intended to be two separate items. Still, the way the introduction is coded cannot ignore the presence of the claim so these items may covary due to this method variance (Brown, 2003; Marsh, 1996). All respecifications were justified by both theory and empirical data (MacCallum, Roznowski, & Necowitz, 1992).

The respecified model was run and had excellent fit. The RMSEA value was smaller than

.08, the CFI and Tucker Lewis Index (TLI) values were ideal at 0.97 and 0.96 respectively, and

the SRMR value was less than 0.10, which is also considered ideal. Compared to the original

four-factor model, the Chi-square difference of fit test supported the preference for the

respecified model ($\Delta\chi^2 = 187.621$, df = 5, $p < .001$).

However, an inspection of the model indicated a serious problem with the double loading

of the body onto both the Structure/Ideas and Evidence factors. Specifically, it resulted in a

negative factor loading from the Ideas/Structure factor to the indicator body (-.159). For this

reason, a final model was run in which body was restricted to only load onto Evidence Use. The

results of this final respecification are reported in Table 11.

Table 11.
*Respecified Writing Model Fit*

	χ^2, DF; p-value	CFI	RMSEA	SRMR	Model comparison
Figure 2c respecification*	285.97; 141; <.001	.97	.070	.031	
Figure 2c second respecification	**289.065; 142; <.001**	**.96**	**.071**	**.031**	**$\Delta\chi^2 = 3.095, \Delta df = 1, p = .08$**

Note. * indicates the model where body was set to load onto two factors. The bolded model constrained
evidence to load onto a single factor—Ideas/Structure.

As seen in Table 11, the fit for the final respecified model was good. The chi-square

difference test also indicated a preference for this as the p-value was non-significant ($\Delta\chi2 =$

3.095, $\Delta df = 1$, p = .08), indicating the preference for the more parsimonious model with a single

factor loading for the body. See Figure 14 for the final model.

Figure 14.
Model of Best Fit for Source-based Argument Writing in History
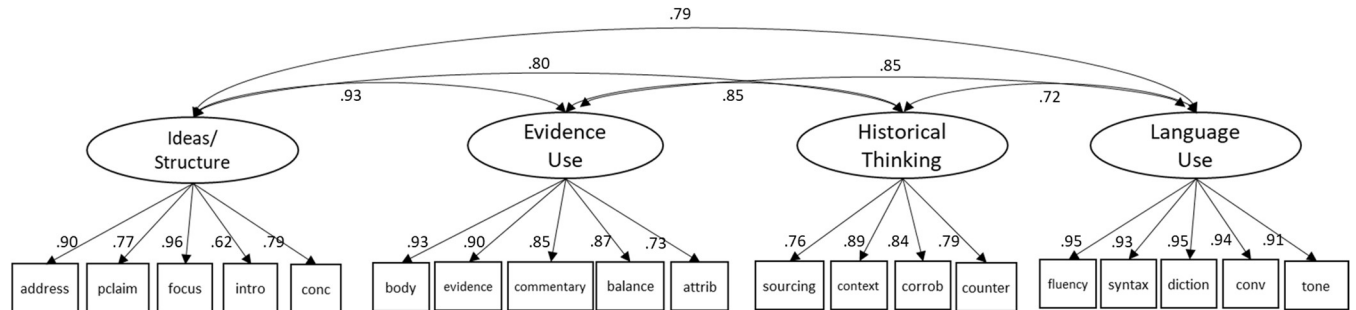
Note. Standardized estimates are reported.

The implication of this loading, which was contrary to the hypothesis, is worth discussing. The quality of evidence, commentary, and a balance of the two are indicators for the latent factor Evidence Use. The evidence and commentary are also the substance, the raw fodder, of the body of an essay that is being organized. Therefore, the respecification does make sense, despite prior hypotheses. Viewing student writing as a complex set of interrelated skills (Steiss et al., 2022) a student's skills with evidence use, providing and explaining how the evidence supports the claim, directly informs how well the body is structured. This finding is aligned with writing pedagogies which utilize a C.E.R. (Claim. Evidence. Reasoning) heuristic for organizing the body, emphasizing the presence of evidence and reasoning to successfully structure the body of an essay.

In the final model, all factor loadings from indicators to their respective latent factors were moderate or strong (.615 ≤ 964), signifying that performance in each analytic item is moderately to strongly related to student proficiency in that specific latent factor. At the same time, all factors were strongly related (.72 ≤ .93). Overall, the results *partially* confirm my hypothesis that SBAW in history would be multidimensional as I rejected the five-factor model for the four-factor model. I also expected Historical Thinking to share less strong correlations with other factors, but the correlations were strong across all factors. This aligns with other research that views historical thinking as a separate dissociable dimension of writing quality in

history, though one strongly correlated with other indicators of writing quality (De la Paz et al., 2017).

Relations to holistic scores

To examine the relations between dimensions of SBAW in history and holistic writing quality, two structural regression models were fit to the data. In the first model, the dimensions of writing were used as latent factors to predict the holistic score. In this model, three of the four latent factors predicted the holistic score: Ideas/Structure ($b = .363$, $SE = .108$, $p < .01$), Evidence Use ($b = .331$, $SE = .143$, $p < .05$), and Historical Thinking ($b = .183$, $SE = .065$, $p < .01$). Contrary to hypotheses, the Language Use factor did not significantly predict the holistic score ($b = .093$, $SE = .108$, $p = .126$). Overall, dimensions of writing quality predicted holistic scores well—the model captured a total of 85% of the variance in the holistic scores.

Next, three controls were added to the structural regression model and were allowed to predict each dimension of writing and the holistic score. See Figure 15.

Figure 15.

Dimensions of Writing Predicting Holistic Scores while Controlling for Sex, EL Status, and HS
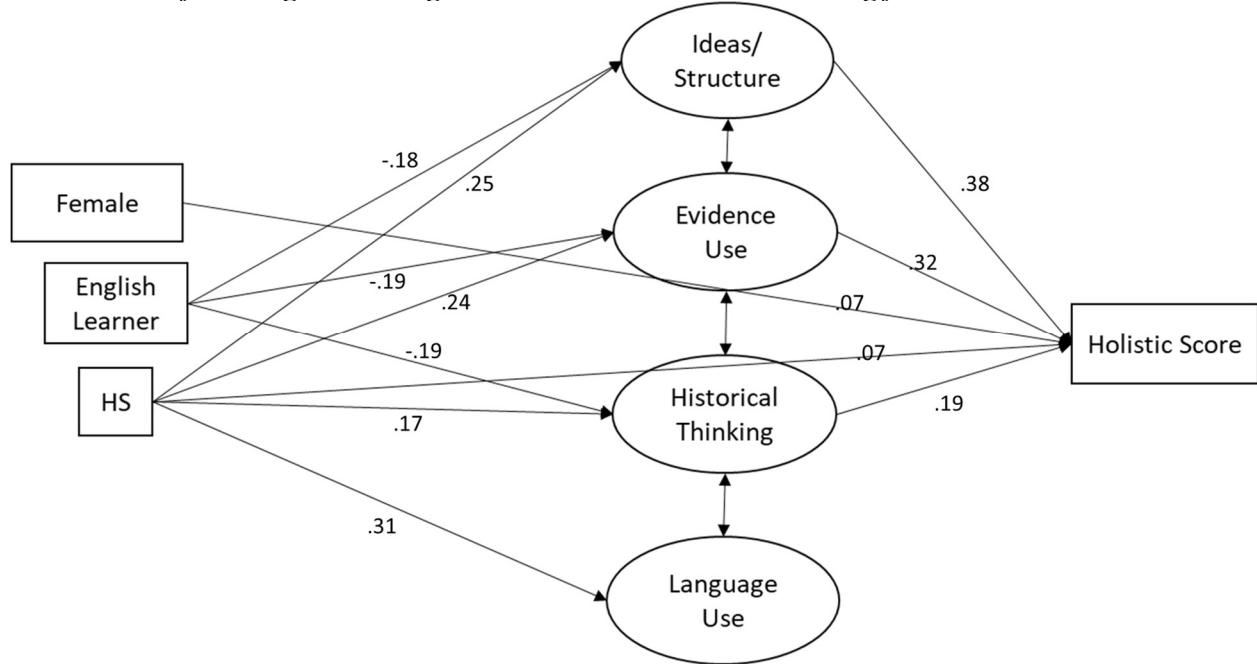
Note. Standardized estimates are reported.

In this model, Ideas/Structure, Evidence Use, and Historical Thinking still significantly predicted holistic scores with a similar magnitude in the model without controls (Ideas/Structure: $b = .38$, $SE = .108$, $p < .01$; Evidence Use: $b = .32$, $SE = .143$, $p < .05$; Historical Thinking: $b = .19$, $SE = .065$, $p < .01$).

Being female had a direct effect on holistic scores, albeit a small one ($b = .07$, $SE = .029$, $p < .05$) and no significant effect on the latent factors. Being in high school, compared with middle school, also had a direct affect on the holistic scores ($b = ..07$, $SE = .030$, $p < .05$) and indirect effects through the Ideas/Structure ($b = .25$), Evidence Use ($b = .24$) and Historical Thinking ($b = .17$) factors.

Being an EL student predicted lower scores in Ideas/Structure ($b = -.18$, $SE = .069$, $p < .01$), Evidence Use ($b = -.19$, $SE = .069$, $p < .01$), and Historical Thinking ($b = -.19$, $SE = .070$, p

< .01) factors, but there was no significant relation between EL status and the Language Use factor. Overall the model explained 86% of the variance in the holistic scores.

Discussion

Writing performance across grade levels and EL Status

Research shows that academic writing is a challenging skill to develop and takes time, especially for ELs (Olson et al., 2020a; Kanno & Cromley, 2015). The present study confirms this claim and extends previous work by examining SBAW in history. As hypothesized, students were challenged with many skills related to evidence use and historical thinking, especially commentary/reasoning, sourcing, corroboration, and presenting and addressing counterarguments. While there were certainly grade level differences for items measuring language use and the presentation of ideas, the lower scores for items related to disciplinary thinking and writing were notable.

In history, it is apparent that many aspects of general argumentation (e.g. providing reasoning to support claims) and disciplinary thinking (e.g. sourcing and addressing counterarguments) are challenging for secondary students (Wineburg, 1991; Goldman et al., 2016). While learning new features of a genre is difficult, this difficulty may be compounded by disciplinary considerations of history. The overall lower scores for historical thinking affirm Wineburg's claim that historical thinking is "an unnatural act" (Wineburg, 1991). Students do not enter classrooms with sophisticated disciplinary reading, thinking, and writing practices to make arguments about the past. Instead, students need explicit instruction, modeling of key practices, and opportunities to practice source-based inquiry to develop disciplinary skills (De La Paz et al., 2017). The differences between sourcing and contextualization also suggest that a

variety of prompts should be utilized to elicit all types of historical reasoning students need to make sense of the past.

The lower scores on reasoning across grade levels suggest students need more high-quality instruction that emphasizes knowledge transformation (Bereiter & Scardamalia, 1987) and higher-order reasoning tasks that ask students to construct meaning from documents, think in discipline-specific ways, and use evidence and their reasoning to make defensible claims about the past (Monte-Sano, 2011, 2012). Providing commentary or reasoning is an important area for development for all students. Research indicates engaging students in the revision of their writing, dialogue with peers about how the evidence supports claims, and the use of mentor texts can help students develop proficiency in providing commentary and reasoning (Graham et al, 2016; Olson et al., 2017). The emphasis on developing ideas and supporting these ideas with evidence and historical thinking is also supported by the structural regression model which shows the significant impact the dimensions of Ideas/Structure, Evidence Use, and Historical Thinking have on overall writing quality.

My original hypothesis that older students would perform better than younger students was only partially correct. Grade level changes were observed across most of the analytic framework items; this shows the capacity for student growth that should be acknowledged and leveraged by educators and researchers to design effective instruction. Student differences were notable across grades 6, 7, and 8 as students appear to move from a knowledge-telling period (Scardamalia, & Bereiter, 1987) toward knowledge transformation and increased reasoning with sources.

However, students in grades 10, 11, and 12, for which we see some skills plateau, need instruction in their zones of proximal development to push them further in SBAW. Specific skills

that seem amenable to targeted instruction include reasoning, sourcing documents, and presenting and addressing counterarguments. Teachers can examine grade level differences presented presently to target skills at developmentally appropriate times. However, future studies are needed to assess the malleability of these skills through instruction.

Findings confirmed the hypothesis that EL students would underperform their peers across most writing skills. This corresponds with research in other subjects and disciplines that argues ELs need writing instruction that focuses on all elements of writing, not just language, especially given the gaps in many elements of writing occurred *outside* the Language dimension (Olson et al., 2020; Fitzgerald, 2017; Short & Fitzsimmons, 2007). While unfamiliarity with the English language hinders EL students as they try to make sense of documents and articulate argumentative claims, this does not necessarily mean instruction focusing on higher-level thinking should be ignored. RFEP students' performance compared with their EO/IFEP peers speaks to the capacity for English learners to grow over time in argument witting.

Dimensions of writing

The stance that all aspects of writing should be targeted during instruction is supported by the factor structure of SBAW in history—all the factors were related and had significant loadings onto analytic framework items. These results were inconsistent with previous studies that find writing quality to be unidimensional (Graham et al., 2005; Olinghouse, 2008) or argue analytic scores are too closely correlated to provide additional information about writing quality (Bang, 2013; Crossley & McNamara, 2010).

Using multiple indicators of quality informed by rigorous human coding, we find dissociable dimensions seen in previous research. For example, we find that language features are a dissociable construct to measure writing (Kim et al., 2014), Evidence Use is a complex and

distinct skill in text-based analytical writing (Wang et al., 2018; Correnti et al., 2020), and historical thinking is a significant component of SBAW in history that is related to overall quality (De La Paz et al., 2017).

Despite our hypothesis that Ideas and Structure would be dissociable dimensions, they were too strongly correlated in the five-factor model and fit well as a single dimension in the four-factor model. This correlation and single dimension fit suggest the Ideas and Structure of writing are too inextricably bound and not separate constructs (Steiss et al., 2022). Yet, combining them into a single construct is at odds with assessment practices that frame the structure of an essay as distinct from its ideas (National Writing Project, 2005, 2010; NREL, 2011).

For example, when attending to the quality of an introduction, a scorer thinks about how the introduction organizes key ideas, such as a claim, that carries ideas throughout the essay in an organized body and a conclusion that restates these main ideas. In this way, the writing's "structure" and "ideas" are interwoven. Similarly, in structuring the body of an essay, writers need skills in selecting, integrating, and analyzing evidence from sources. The reasoning should also connect back to the main claim. Therefore, the structure of a body is connected to the generation of ideas and skills in evidence use.

In regards to writing evaluation and instruction, classroom teachers, may not find a 24-item method economical and may instead assess students' Structure/Ideas, Evidence Use, Historical Thinking, and Language Use as separate items of writing. For example, a teacher may want to target students' skills in Evidence Use after noticing a lack of commentary/reasoning. Given the strong correlations between items within latent dimensions, focusing on one specific skill, like providing reasoning, seems useful and economical to improve student writing within

that dimension. Additionally, the strong correlations between factors emphasize a holistic approach to instruction where all elements are valued.

Dimensions of writing predicting holistic scores

The structural regression model shows that the analytic framework and human coding can reliably describe overall writing quality while attending to the relative contributions of its dissociable dimensions. While three dimensions of writing quality in this genre were significantly related to the holistic score, the moderate relations of Ideas/Structure and Evidence Use to the holistic score speaks to the importance of skills like presenting a clear and arguable claim, integrating evidence, presenting commentary, and balancing evidence, commentary, and summary in SBAW in history. The significant contribution of Historical Thinking is also noteworthy and given the low scores for the skill, its contributions to holistic scores might change in a sample with higher average scores in this dimension. The results show that students who can use evidence to substantiate claims and ideas, select, integrate, and comment on the evidence, and engage in skills like contextualization are better positioned for successful SBAW in history. Given the lower scores observed for these elements, educators should be encouraged and supported in targeting these areas to improve holistic writing scores.

Contrary to my initial hypothesis, the Language Use factor was not related to the holistic score. This should be interpreted alongside the strong correlations between Language Use and other factors, which indicate it is related to other dimensions of writing which predict holistic quality. Still, ELs' lower performance in holistic writing only occurred through the Ideas/Structure, Evidence Use, and Historical Thinking factors—not Language Use. This non-significant influence indicates that a curricular approach focused on vocabulary and grammar while ignoring the higher-order reasoning skills needed to engage in commentary and

argumentation is misguided and can exacerbate current opportunity gaps (Applebee et al., 2003). Instead, these results suggest comprehensive, literacy-rich curricula that emphasize disciplinary reasoning is the key to improved writing for ELs and other students (Goldman et al., 2016; Olson et al., 2017; Olson et al., 2020).

Limitations and Conclusion

A key limitation of this sample is the low number of teachers represented at each grade level which makes the students less representative than a sample across many classrooms. The relatively low scores for grade 10 students indicate this issue may affect results and a broader sample should be used in subsequent studies. Additionally, the lack of meaningful covariates to explain classroom level differences exacerbates this issue and constrains additional analyses. For example, students' designated RFEPS could not be added to the structural regression model, even though their writing may be different from students designated EO/IFEP (Olson et al., 2017)

Another limitation is that scored writing samples come from only a single prompt. As noted previously, this prompt may prioritize certain thinking skills over others—namely a preference for contextualization over sourcing. As writing prompts influence student writing performance (De La Paz & Monte-Sano, 2012), more writing prompts should be used to capture the construct of interest: students' SBAW in history. The factor structure in the present study might change if different prompts are used that emphasize different skills. Similarly, the low scores for items related to historical thinking may also influence factor structure and the contributions of distinct factors to holistic scores. Analyses should also be completed with students scoring higher in these items to understand the contributions of Historical Thinking to holistic scores in a higher scoring sample of students.

Overall, an analytic picture of students' SBAW in history shows the challenges students face reasoning with historical sources and arguing using disciplinary literacy. While all elements of writing are important, developmental differences that exist should be acknowledged in designing writing instruction. Such an analytical approach allows researchers and educators to make informed decisions about what skills to target as students develop. Present findings contribute to a growing body of research that affirms English learners' path to writing proficiency features instruction that emphasizes the construction of knowledge with evidence and interpretative, higher-order thinking.

Study 2: What are the features of secondary students' epistemologies in history?

Purpose

Given the need to better understand the epistemologies of secondary students, the second study examined the epistemologies of students in grades 6-12. The study was guided by the following research question: *What are the features of secondary students' epistemologies in history?* The findings from Study 2 intended to identify key dimensions of students' epistemologies and use them as latent factors in Study 3 which examines the relations between epistemology and writing.

Using a developmental approach to measure epistemology, I hypothesized that students would have relatively naive epistemologies, with many students endorsing absolutist or multiplist views of knowledge in history, and more high schoolers endorsing evaluativist views of knowledge. Using a dimensional approach, I expected students to have moderate to low beliefs in knowledge being Complex or Tentative in history given common approaches to teaching history as the transmission of fixed and static knowledge (Bain, 2005, 2006; Goldman et al., 2016; Monte-Sano, 2010; Reisman, 2012). I had no a priori hypothesis about

79

Complex/Tentative being a single factor or two factors given unclear findings from prior research (Strømsø & Braten, 2010; Wiley et al., 2020). Finally, I expected students to show a moderate belief in the value of integration to support claims.

Method

Participants

Participants in the second study included 75 students from Study 1 and an additional 185 students (N=256). The 75 students were those who completed the epistemology survey and were selected for analytic coding in Study 1. The 185 students were the remaining students who completed the epistemology survey. Though all teachers in the larger field trial administered the writing tasks to their classrooms, not all teachers administered the epistemology survey measure discussed next, despite directions to do so. Given the Covid-19 pandemic teachers and students were facing and the need to administer the two-day writing task and an additional survey measuring self-efficacy towards writing, it is likely many teachers in the field trial decided to omit or simply forgot to administer the survey. Additional efforts were taken to increase survey administration such as using district leads to contact teachers. For this study, only students who were in classrooms where the survey was administered are included. The following table shows the grade level breakdown for the 256 students in the sample,

Table 12.
Epistemology Survey Responses by Grade Level

	All Grades	Grade 6	Grade 7	Grade 8	Grade 10	Grade 11	Grade 12
Responses	256	37	22	60	17	26	48

Measures

Epistemology Scales

In designing a measure of students' epistemologies in history, I first reviewed extant

literature related to the measurement of epistemology in history and other domains. The dimensional approach (placing students on a spectrum across multiple dimensions) and the developmental approach (placing students into one of three increasingly sophisticated epistemological categories) were both represented in studies taking place in history classrooms and other domains (Bråten et al., 2014; Buehl & Alexander, 2005; Maggioni et al., 2009; Wiley et al., 2020). Consequently, both approaches were used in the study.

Dimensionality measure

Measuring students' beliefs about knowledge across two dimensions—Simple/Complex and Certain/Tentative can be found in early descriptions of students' epistemologies in history (Voss et al., 1995; 1998). Most recently, Wiley and colleagues (2020) developed an epistemic beliefs scale to measure secondary and undergraduate students' epistemologies in history. They followed the dimensional approach used in multiple source-based reasoning tasks in other domains (see Bråten & Strømsø, 2009) and found epistemic beliefs to independently predict the reasoning and writing performance of high school students. They also developed an Integration subscale to measure the value of integrating evidence when learning from multiple documents in history. This subscale reflects activities historians engage in during inquiry as they integrate evidence from multiple sources and perspectives to validate claims. However, distinct from previous research, they found that the Simplex/Complex and Certain/Tentative dimensions were one dimension using principal components analysis (PCA).

Wiley and colleagues' (2020) scale was reliable and valid across multiple experiments and subgroups of students and was, therefore, modified in the present study. The single factor was investigated by adapting items in their scale so sufficient items were available to test the dimensionality of the Simple/Complex and Certain/Tentative factors in confirmatory factor

analysis, a more theory-driven approach than PCA.

Further review of extant literature led to the inclusion of 10 items for each potential dimension of epistemology: Simple/Complex, Certain/Tentative, and Integration (Bråten & Strømsø, 2009; Bråten et al., 2014; Buehl, & Alexander, 2005; Wiley et al., 2020). The first two potential dimensions reflect beliefs *about* the nature of knowledge, with more sophisticated beliefs viewing knowledge as *complex* (interrelated theories and facts and multiple causes or explanations) and tentative (explanations and truths can be revised or changed with new information). The third dimension reflects beliefs about how knowledge claims should be justified, with more sophisticated beliefs prioritizing the justification of claims by integrating evidence from multiple sources.

From this point, a team engaged in item reduction and modification of items by soliciting feedback from four subject-matter experts in the field of writing beliefs and history. Items were also modified for language to ensure the scale was developmentally appropriate for the sample, which featured middle school students and a large population of English learners. Eventually, a 14-item scale was produced, which can be seen in Table 13.

Table 13.
Items Measuring Epistemology Using a Dimensional Approach

Hypothesized Dimension	Item
Simple/Complex	Historical events are due to a single cause*
Simple/Complex	The best explanations in history stick to one major cause of an event*
Simple/Complex	There is only one good historical explanation that can be written from a set of facts*
Simple/Complex	To understand the causes of historical events, you should consider many perspectives and explanations
Simple/Complex	History is more like a set of facts than competing interpretations of facts*

Certain/Tentative	Accepted explanations in history can be proven wrong
Certain/Tentative	Historical explanations should change when new information becomes available
Certain/Tentative	What is considered to be true about history today may be considered to be false tomorrow
Certain/Tentative	I think simple and clear explanations about historical events are better, even if they are not the most accurate*
Integration	When you read about history, you should trust explanations based on evidence
Integration	To understand the causes of historical events, you need to connect evidence using reasoning
Integration	A good way to decide if what you read is accurate is to compare multiple sources
Integration	When you read about something new in history, you should compare it to what you have already learned to see if the new information is trustworthy
Integration	If you read something in a textbook for history, you can be sure it's true*

Note. Items marked with an asterisk were reverse coded so responses were all positively oriented across dimensions; the highest response represents a view of knowledge of complex, tentative, and justified through the integration of evidence

Student responses were collected using a sliding scale from 1 to 10, with 1 representing that they strongly disagree with the statement and 10 indicating they strongly agree with the statement.

Developmental measure

Given the use of scales taking a developmental view of epistemology across many domains (e.g. science, history), this type of measure was also included in the study. Items on this scale intended to categorize students into different stages of epistemological maturity: Absolutists see knowledge as a concrete and stable set of facts and assert there is one correct view of the world; multiplists see multiple truths can exist simultaneously; and evaluativists believe knowledge is produced by reconciling multiple interpretations with evidence and reason

(Barzilai & Weinstock, 2015; Kuhn, 2019). The measure was adapted from Barzilai and Eshet-Alkalai (2015). Modifications to their questions, which focus on scientific views about climate change, were made by referencing Maggioni et al. (2009) who developed a similar scale in history. With the help of experts in the field, we balanced the ease of use of the measure students in the sample and eliminated redundant questions. Eventually, a 4-question scale was developed. The scale can be seen in Figure 16.

Figure 16.
Items Measuring Epistemology Using a Developmental Approach

*Indicate which of the following responses is **best**?*
1. How should you evaluate explanations about historical events?
    a. Check if the explanation reports exact facts and not opinions.
    b. Check if the explanation matches your view of events.
    c. Check if the explanation helps improve your understanding of what is known about the event.
2. What should knowledge about historical events be based on?
    a. Only on the facts.
    b. Mainly on personal points of view.
    c. Mainly on interpretations of facts and points of view.
3. Is there truth about the causes and effects of historical events?
    a. There is truth. If it is not known it is important to find it out.
    b. There is no single truth; there are just different opinions and interpretations
    c. There are different interpretations of events, but some interpretations are more accurate than others.
4. Is there an answer to the question of what led to the success of a certain movement?
    a. There is one right answer.
    b. It is impossible to know the right answer.
    c. There may be multiple explanations, but some are better than others.

In Figure 16, option A represents an absolutist response, option B represents a multiplist response, and option C represents an evaluativist response. As the scale requires students to choose one of three multiple-choice questions, students can be placed into categories by having each response scored based on complexity, with one point for the corresponding absolutist response, two points for the corresponding multiplist response, and three points for the corresponding evaluativist response. Researchers have also categorized students based on their patterns of responses (Kuhn et al., 2000). This latter method of categorizing responses was used in the present study as it more specifically attends to the specific nature of each student's responses, as opposed to simply adding responses together. For example, a student indicating an

absolutist response 3 times and an evaluativist response one time, would reasonably be categorized as an "absolutist." However, simply adding responses might inaccurately identify the student as a multiplist.

Both epistemology measures were administered by the classroom teacher as a single student survey intended to measure students' beliefs about knowledge and history. The survey was taken using Qualtrics, with all survey items randomized for the sliding scale questions and all answer options randomized for the multiple-choice questions.

Analytic approach

To answer the research question, *What are the features of secondary students' epistemologies in history,* data from the dimensional version of the epistemology survey and CFA were used to examine the dimensions of students' epistemologies in history. These dimensions were intended to be used as latent factors in Study 3 to describe the relations between epistemology and writing.

I also intended to use the developmental measure to describe students' epistemological development across grade levels. However, several issues made this measure unacceptable for further analyses. First, responses to each question were too frequently in the evaluativist category given what prior research would indicate is plausible. There appeared to be an acquiescence bias with students preferring the longest and more detailed responses, which also represented the evaluativist level of epistemological development (see Table 14).

Table 14.
Student Responses by Category of Epistemological Development

Category	Frequency	Percentage of Students
Absolutist	62	24.4%
Multiplist	77	30.2%
Evaluativist	116	45.4%

Additionally, after categorizing student responses across various levels of epistemology, this variable was not significantly related to other variables in students' writing or 13 of the 14 items in the dimensional epistemology measure. The variable also had unsatisfactory distributional properties, with responses indicating many more students were evaluativists than prior research would suggest is plausible. For these reasons, the use of the dimensional view was not used.

Returning to the dimensional measure, before conducting CFA, variables were examined for their univariate and bivariate distributions and multivariate normality (see Table 15)

Table 15.
Items Measuring Dimensions of Students' Epistemology

	Mean	SD	Minimum	Maximum	Skewness	Kurtosis
Events have multiple causes*	6.71	2.43	1.00	10.00	-0.36	2.22
Best explanations list many causes*	5.33	2.27	1.00	10.00	0.07	2.50
Multiple explanations can be written*	6.42	2.36	1.00	10.00	-0.26	2.38
Need many perspectives to understand events	7.98	1.95	1.00	10.00	-0.84	3.06
History as competing interpretations of facts*	5.05	2.04	1.00	10.00	-0.09	2.90
Explanations can be proven wrong	6.80	2.28	1.00	10.00	-0.21	2.23
Explanations should change w/ new info	6.67	2.37	1.00	10.00	-0.29	2.39
True today - False tomorrow	6.57	2.54	1.00	10.00	-0.29	2.23
Complex and accurate explanations preferred*	5.34	2.21	1.00	10.00	0.25	2.62
Trust evidence-based explanations	7.51	2.08	1.00	10.00	-0.63	2.89
To understand causes need evidence and reasoning	8.00	2.02	2.00	10.00	-0.90	3.11
Compare multiple sources for accuracy	8.03	2.02	1.00	10.00	-0.86	3.05
Compare sources when verifying new info	7.10	2.10	1.00	10.00	-0.36	2.58
Textbooks not necessarily true*	4.99	2.42	1.00	10.00	0.20	2.45

Note. Responses ranged from 1-10 and were recorded by sliding a bar between strongly disagree and strongly agree. An asterisk indicates responses that have been reverse coded. Refer to Table 13 for the original wording of the item.

An examination of the uni- and multivariate properties of variables indicated all variables were normal and adequate for engaging in CFA. Items that were originally worded with negative language did have lower overall means than other items. For example, complex1 (negatively worded) had a mean of 6.71 compared to complex4 which had a mean of 7.98 and was positively oriented. Bivariate correlations between variables are also presented in Table 16.

Table 16.
Bivariate Correlations for Items in Dimensional Epistemology Scale

	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1. complex1	1													
2. complex2	.44*	1												
3. complex3	.39*	.37*	1											
4. complex4	.03	-.06	.08	1										
5. complex5	.20*	.35*	.36*	-.20*	1									
6. tent1	.05	.05	.09	.25*	.01	1								
7. tent2	-.01	-.19*	-.11	.29*	-.06	.23*	1							
8. tent3	-.01	.03	.10	.23*	.00	.36*	.34*	1						
9. tent4	.22*	.28*	.28*	-.04	.26*	-.14*	-.05	-.07	1					
10. just1	.02	-.24*	-.05	.32*	-.20*	.15*	.16*	-.01	-.15*	1				
11. just2	-.05	-.13*	.03	.45*	-.17*	.24*	.27*	.15*	-.06	.36*	1			
12. just3	0.13	-.08	.04	.39*	-.10	.26*	.22*	.25*	-.20*	.26*	.38*	1		
13. just4	.01	-.22*	-.12	.42*	-.26*	.21*	.19*	.13*	-.07	.25*	.36*	.15*	1	
14. just5	.20*	.28*	.30*	-.08	.25*	.03	-.04	.16*	.31*	-.31*	-.09	-.14*	-.06	1

Note. Asterisk* indicates $p < .05$

While I expected all variables to be somewhat related and for hypothesized dimensions to be more strongly correlated, there were only weak or significant correlations across all variables and correlation patterns did correspond to hypothesized dimensions—many variables showed stronger correlations with variables from different hypothesized dimensions than variables within the same hypothesized dimension. For example, complex3 (originally worded: "There is only one good historical explanation that can be written from a set of facts") has an insignificant correlation with complex4 ("To understand the causes of historical events, you should consider many perspectives and explanations") at .08, but is significantly correlated with tent4 at .22 ("I

think simple and clear explanations about historical events are better, even if they are not most accurate").

I conducted a close analysis of items, noting which were significantly related and which were not. Examining pairs of items, I found that tent2 and complex3 appeared to ask a very similar question: Tent2, asks if "historical explanations should change when new information becomes available" while complex3, which is a negatively worded item, asks if "there is only one good historical explanation that can be written from a set of facts." Both suggest there are multiple possible historical explanations and it is unlikely someone's responses to these items would not correlate. On further examination, it appeared the only difference between these items was item orientation. The correlations suggested there were response patterns based on whether items were positively or negatively oriented items. These patterns appeared to be unrelated to the hypothesized dimensions of epistemology.

Presently, I define negatively oriented items as those that, when endorsed, indicate a lower amount of a measured trait—in other words, items that have an inverse relation to the construct of interest (van Sonderen et al., 2013; Bandalos, 2018). For example, a "strongly agree" response to the complex1, "Historical events are due to a single cause*," indicates low levels of Complexity beliefs. Positively oriented items indicate that, when endorsed, a respondent has higher levels of that construct (for example, high beliefs that knowledge is complex or tentative). To illustrate, tent2, "Historical explanations should change when new information becomes available," is *positively* oriented, with a response of 10, strong agreement, indicating a belief that knowledge is tentative.

Returning to the correlation matrix, *all* correlated items in the epistemology scale were those sharing item orientation. Additionally, it appeared that extant literature used scales to

measure dimensions of epistemology that were possibly confided by item orientation. If this was the case, it suggests patterns of responses based on item orientation may generate independent factors that represent "artifact[s] of the language of item[s] rather than underlying construct[s]" (Deuber et al., 2021, p.2). Given the potential for item orientation to confound factor structure, its influence was investigated (Deemer & Minke, 1999; Johanson & Osborn, 2000; Ibrahim 2001; Weems et al., 2003; DiStefano & Motl, 2006).

Findings

Investigating Item Orientation

First, I retroactively examined my own decisions regarding item orientation and apparent decisions made by previous research. Second, I added another model to test using CFA in which items were set to load onto two factors reflecting positive and negative item orientation (See Figure 17). This model uses item orientation as the best explanation for the common variance of items. Because CFA is theory-driven, the theory for factors reflecting item orientation—Positive and Negative—needed to be established. Therefore, I first discuss the construction of my scale and adaptions made by previous researchers, before conducting the initially planned CFA with the additional model that assesses the influence of item orientation on factor structure.

I first examined Wiley et al.'s (2020) scale, which was modified for the present study. Their first factor, titled Simple/Certain, features items that are uniformly **negative** in their orientation (e.g. "Historical explanations should **not** change in light of new information"). The second factor, titled Integration, features items that are positively oriented (e.g. "To understand the causes of historical events, you need to connect evidence using reasoning")[1]

---

[1]Others might argue that items are positively oriented because the scale is titled "Certain/Simple." High responses indicate high levels of simple or certain beliefs. However, previous research is clear that Complex and Tentative views of knowledge are more adaptive and more frequently use these terms. Further, the 49-item Topic-Specific Belief Questionnaire, from which Wiley et al. (2020) and others construct their scales name the dimensions

When developing the present scale, I adapted some items to test if the Simplex/Certain factor was one or two dimensions as previously validated scales conceptualized Complex and Tentative views of knowledge as distinct and dissociable dimensions (Bråten & Strømsø, 2009; Strømsø et al., 2008). I also adapted items for readability and to ensure there were some positively and negatively oriented items in each scale to reduce acquiescence bias or satisficing responses (Barnette; 2000). To illustrate, the following table shows 1) items originally used by Wiley and colleagues (2020), 2) the dimension these items loaded onto in their research, 3) changes I made to these items, and 4) the dimensions they loaded onto after changing the item's orientation.

Table 17.
Change in Epistemology Items and Dimensions

Wiley et al. (2020) item	Dimension	Present study item	Dimension
There is only one good historical explanation that can be written from a set of facts.	Component 1 (Simple/certain)	There is only one good historical explanation that can be written from a set of facts.	Negative
Most historical events are due to a single cause.	Component 1 (Simple/certain)	Historical events are due to a single cause.	Negative
Good explanations in history are always indisputable.	Component 1 (Simple/certain)	Accepted explanations in history can be proven wrong	**Positive**
Historical explanations should not change in light of new information.	Component 1 (Simple/certain)	Historical explanations should change when new information becomes available.	**Positive**
You can be certain that historical explanations are true.	Component 1 (Simple/certain)	What is considered to be true about history today may be considered to be false tomorrow.	**Positive**

While all of Wiley et al.'s (2020) items measuring Simple/Certain views of knowledge were negative, when items were changed to a positive orientation in the present study, they

Complex and Tentative and feature positively and negatively oriented items for these dimensions. Wiley and colleagues (2020), through PCA, essentially remove positively oriented items from the Complex and Tentative dimensions and rename the scale Simple/Certain. The following analyses should make clear that changes in item orientation were made to create this resulting factor structure.

loaded onto a distinct dimension as seen in rows 3-5 in Table 17. Row 4 offers a particularly salient example as "historical explanations **should not** change in light of new information" is changed to "historical explanations **should** change when new information becomes available" and subsequently loads onto a distinct dimension (despite the content on the item being virtually identical). I interpret this as initial evidence that changing item orientation can have consequential changes on the factor structure of constructs in a way that undermines claims about dimensionality. Next, I examined previously validated scales that Wiley and colleagues (2020) used to create their present scale.

Like Wiley et al., (2020), Bråten & Strømsø (2009) used the Topic-Specific Epistemic Beliefs Questionnaire (TSEBQ) to create an epistemic beliefs scale to examine the dimensions of student epistemology related to climate change. Their scale has been utilized or modified in hundreds of subsequent studies. Bråten & Strømsø (2009) posit that Simplicity and Certainty are two distinct dimensions of epistemic beliefs about climate change. However, items in the Certainty dimension are positively oriented, with questions like: "Theories about climate change can be disproved at any time" and "What is considered to be certain knowledge about climate research today may be considered to be false tomorrow." Both these questions imply knowledge can change.

The "dissociable" Simplicity dimension features negatively oriented questions that resemble the content of items in the Certainty dimension: "Knowledge about climate change is indisputable." This item, when reverse coded, also implies knowledge can change and should plausibly load onto the same dimension as the aforementioned Certainty items. Either this conjecture is incorrect, or item orientation explains the loading of the item. Further analysis, I argue, supports the latter hypothesis.

91

The respective items for Certainty and Simplicity dimensions are enumerated below, with the item orientations labeled as follows: positive indicates agreement with a sophisticated epistemology (Tentative and Complex) and negative agreement aligns with a less sophisticated epistemology. This is in line with the original TSEBQ and with the practice of researchers who reverse code these items so that sophisticated or adaptive epistemic beliefs positively predict reasoning outcomes (Bråten, & Strømsø, 2009; Yli-Panula et al., 2021).

Table 18.
Items Measuring Epistemology with Uniform Item Orientation within Dimensions

Certainty of knowledge	Item Orientation	Simplicity of knowledge	Item Orientation
What is considered to be certain knowledge about climate today may be considered to be false tomorrow.	Positive	The knowledge about climate problems is indisputable.	Negative
Theories about climate can be disproved at any time.	Positive	With respect to knowledge about climate, there are seldom connections among different issues.	Negative
Certain knowledge about climate is rare.	Positive	Knowledge about climate is primarily characterized by a large amount of detailed information.	Negative
The results of climate research are preliminary.	Positive	Within climate research, accurate knowledge about details is the most important.	Negative
The knowledge about issues concerning climate is constantly changing.	Positive	Within climate research, various theories about the same will make things unnecessarily complicated.	Negative
Problems within climate research do not have any clear and unambiguous solution.	Positive	There is really no method I can use to decide whether claims in texts about issues concerning climate can be trusted.	Negative

Note. While the scale uses uniform item orientation within dimensions, the scale also uses polarized item orientation across dimensions (one subscale is negatively oriented and one subscale is positively oriented).

Although the above table does not conclusively show item orientation creates factor structure, it is suggestive. Bråten and Strømsø (2009) and Wiley and colleagues (2020) also use Principal Components Analysis (PCA), not CFA, which is sensitive to capturing latent factors in the data that may be "methods" factors (i.e. sharing item orientation) and will parcel these out

(Brown, 2003; Marsh, 1996;). They also report reducing items to by examining fit statistics and coefficient alpha, but not necessarily considering patterns in item orientation.

At this point, I considered similar patterns of uniform item orientation in the extant literature as evidence that we should consider item orientation to be a key contributor to factor structure that must be addressed as a check for robustness. One way to do this would be by to change the item orientation of items used in previously validated scales (Bråten & Strømsø, 2009; Ferguson & Bråten 2013) and then examine subsequent changes in factor structure. Conveniently, Wiley et al. (2020) did just this.

In their article, Wiley and colleagues (2020) describe selecting items from the larger TSEBQ (Bråten & Strømsø, 2009) and combining them with items from Voss (1998) to make them history oriented. They also describe modifying items that had confusing phrases or difficult vocabulary. Notably, they do not mention changing item orientation, though it is apparent this occurred and arguably in a way that made items *less* clear. Also, they removed several items from the 49-item inventory that are positively oriented. They do not mention eliminating positively oriented items explicitly, but these items can be found in other research using the TSEBQ (Bråten, & Strømsø, 2009; Bråten, & Strømsø, 2009; Yli-Panula et al., 2021).

In Table 19, I show the notable changes I observed from the version of the TSEBQ used by Bråten, and Strømsø (2009) to the version used by Wiley and colleagues (2020). I have replaced *climate change* with *topic*, which is then replaced with *history* by Wiley and colleagues. Overall, the table shows the effect of changing item orientation on factor structure. The first three items, already negatively oriented, remain so. The next three items are rewritten from positive to negative item orientation and consequently load onto the same factor as the first three

items despite previous claims that Certainty and Simplicity were distinct dimensions of

epistemology.

Table 19.
Perceived Modifications Made by Wiley et al., 2020 from Extant Research

Items Grouped by Original Dimension	Modification	Modified Item under Simple/Certain Factor
[Negatively oriented] Simplicity of Knowledge Dimension		*[Negatively oriented] Component (Simple/Certain)*
Within *topic*, various theories about the same will make things unnecessarily complicated.	Change topic →	The best explanations in *history* are those that stick just to the one major cause that most directly leads to the event.
Knowledge about *topic* is indisputable.	Change topic →	Good explanations in *history* are always indisputable.
With respect to *topic*, there are seldom connections among different issues.	Change topic →	There is only one good historical explanation that can be written from a set of facts.
[Positively oriented] Certainty of Knowledge Dimension		
What is considered to be certain knowledge about *topic* today may be considered to be false tomorrow.	Change topic, orientation, and dimension →	You can be certain that *historical* explanations are true.
Theories about *topic* can be disproved at any time.*	Change topic, orientation, and dimension →	Historical explanations should not change in light of new information.
Problems within *topic* research do not have any clear and unambiguous solution.	Change topic, orientation, and dimension →	Most historical events are due to a single cause.

Note. *This item is also similar to "The knowledge about *topic* is constantly changing." If this was the adapted item, there is also a change in item orientation and subsequent change in factor structure. The item is also similar to "Knowledge about *topic* is indisputable" which appears in the negatively oriented Simplicity dimension. One might argue there is a substantive difference between something being indisputable versus not being able to be disproved at any time (or knowledge and theory are different) and, therefore, these are indicators of substantively different dimensions of epistemology. However, I see this argument as weak on its own and not robust to critiques of item orientation. One item reoriented would result in the following two questions loading onto different dimensions of epistemology: "Knowledge about *topic* can be disproved" and "Theories about *topic* can be disputed."

Two points are worth considering: First, item orientation is the most salient change in

items made by Wiley et al. (2020). Second, though they claim to make changes to simplify

language, many of the changes are more confusing than the original. For example, "Historical

explanations should not change in light of new information" uses direct negation, which researchers argue should be avoided (Foddy, 1993, Deuber et al., 2021; Gnambs & Schroeders, 2020). One might have simply written: "Explanations about a *topic* should change with new information." This writing would be clear but would likely have a poor factor loading if item orientation matters.

Further, Bråten & Strømsø (2009) ask similar questions about the changing nature of knowledge, but use positively oriented items: "The results of climate research are preliminary" and "What is considered to be certain knowledge about climate today may be considered to be false tomorrow." This second question, I also ask in my own scale about history and find it only correlates with similarly oriented items. This differs from both groups of prior researchers who find it fits somewhere else due to different patterns of item orientation.

Finally, I return to my modifications as evidence of the influence of item orientation on factor structure. I used three items from Wiley et al., (2020) and retained the negative item orientation. These items load onto the Negative factor (discussed in the next section): 1) The best explanations in history stick to one major cause of an event, 2) Historical events are due to a single cause, and 3) There is only one good historical explanation that can be written from a set of facts. The following Table shows modifications I made to other three items.

Table 20.
Modifications to Items Measuring Epistemology

Wiley et al., (2020)		Present study
You can be certain that historical explanations are true.	→	What is considered to be true about history today may be considered to be false tomorrow
Good explanations in history are always indisputable.	→	Accepted explanations in history can be proven wrong.
Historical explanations should not change in light of new information.	→	Historical explanations should change when new information becomes available

Note. Changes made resulted in items no longer being correlated with other, non-changed items.

The table illustrates changes in item orientation, but not content, as a mechanism for the resulting change in correlations and (discussed next) the superior fit of the two-factor Positive and Negative model.

Finally, items used in the present study *not* coming from Wiley et al., (2020) also followed this pattern of being correlated with other items of equivalent orientation. The next table shows these items, their hypothesized dimension, and eventual factor loading on the preferred two-factor model.

Table 21.
New Items Measuring Epistemology

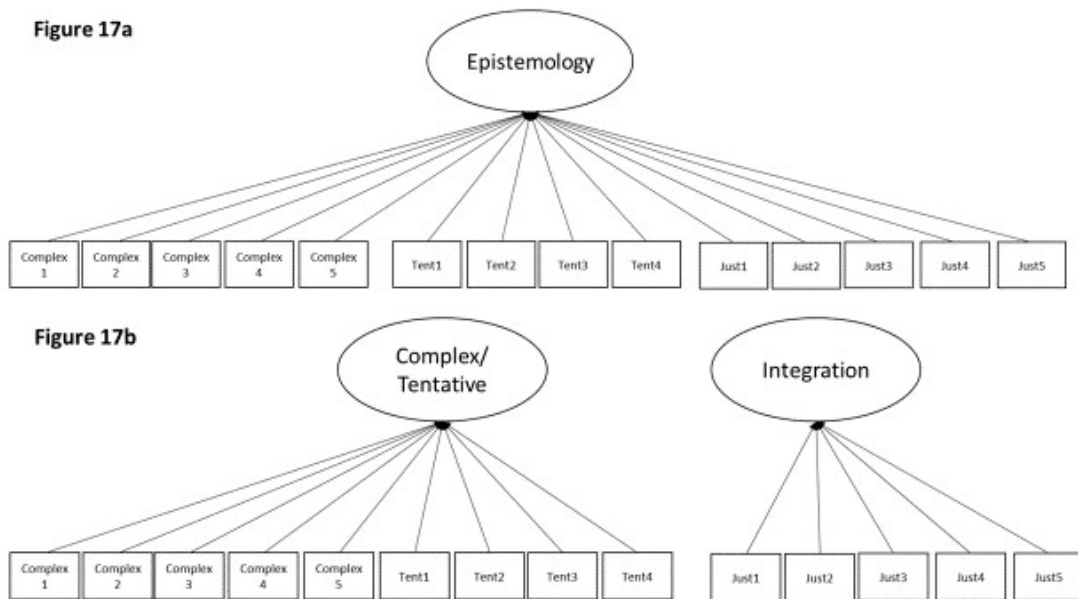| Hypothesized Dimension	Item		Dimension
Complex	To understand the causes of historical events, you should consider many perspectives and explanations	→	Positive
Complex	History is more like a set of facts than competing interpretations of facts	→	Negative
Tentative	What is considered to be true about history today may be considered to be false tomorrow.	→	Positive
Tentative	I think simple and clear explanations about historical events are better, even if they are not the most accurate	→	Negative

Note. These items were originally added to test whether Certain/Simple (termed Complex/ Tentative presently), was a single dimension. The results suggest it is multidimensional, albeit not in a way that prior research argues.
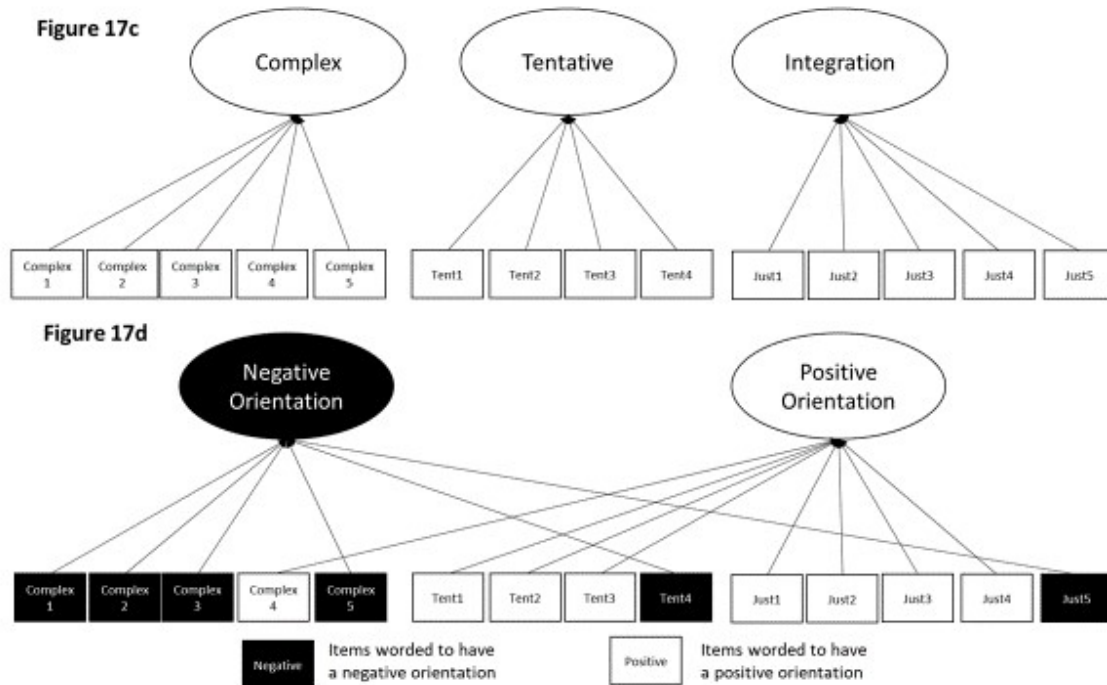
Again, we see the influence of item orientation on student response patterns, correlations, and [eventually] factor structure. A final point to note is that items in the Integration subscale were positively oriented and, therefore, more strongly related to items that were positively oriented from other hypothesized dimensions. Now that the theory for testing an alternative positive/negative model of epistemology has been established, I present the results of the CFA and discuss the implications of this study.

Confirmatory factor analysis

Confirmatory factor analysis using Mplus 8.4 (Muthén & Muthén, 2017) was used to examine the dimensionality of students' epistemology. Four competing alternative confirmatory factor models shown in Figure 17 were fit to the data. The first model (Figure 17a) tested a unidimensional model, the baseline model, where students' epistemology is a single construct that captures all the items in the epistemology. The second model tests a two-factor model similar to Wiley and colleagues (2020) with Simple/Certain and Integration representing two dissociable but related dimensions of students' epistemology measure (Figure 17b). The third model tests a three-factor model (Figure 17c) with Complex, Tentative, and Integration representing three dissociable but related dimensions of students' epistemology. The final model tests a two-factor model where one factor represents the method variance of positive item orientation and another factor represents the method variance attributable to negative item orientation. Model comparisons were conducted using chi-square differences, given that three of these models were nested.

Figure 17.
Alternative models of students' epistemology

Figure 17c

Figure 17d

Items worded to have a negative orientation

Items worded to have a positive orientation

Model fits are reported in Table 22.

Table 22.

Model Fits of Confirmatory Factor Analysis shown in Figure 17

	χ^2, DF; p-value	CFI	RMSEA	SRMR	Model comparison
Figure 17a	417.022; 78; <.001	.460	.130	.144	--
Figure 17b	395.120; 78; <.001	.490	.126	.124	--
Figure 17c	347.540; 77; <.001	.570	.117	.124	4c vs. 4b
					$\Delta\chi^2 = 47.58, \Delta df = 1, p < .001$
Figure 17d	**217.250; 78; <.001**	**.778**	**.084**	**.087**	**4c vs. 4d**
					$\Delta\chi^2 = 130.29, \Delta df = 1, p < .001$

The first three models all had poor fit. The final model (Figure 17d) had slightly better fit, but overall had marginally poor fit. RMSEA was .084, with values below .08 considered ideal, CFI and TLI were .778. with equal or great than .95 considered ideal and .90 acceptable, and SRMR was .087 which is considered acceptable (Hu & Bentler, 1999; Kline, 2015). With one fit statistic acceptable and chi-square difference test suggesting this model had superior fit to the next best-fitting model, Figure 17d was still the best-fitting model.

Factor loadings from positively or negatively oriented indicators significantly loaded onto their respective factors (.326 ≤ .657), but there was a large amount of residual variance or error for each indicator (.568 ≤ .894). This suggests some variance in the item is not being accounted for by the factor. Additionally, the two factors were weakly and negatively correlated (see Figure 18).

Figure 18.
Best Fitting Model of Epistemology
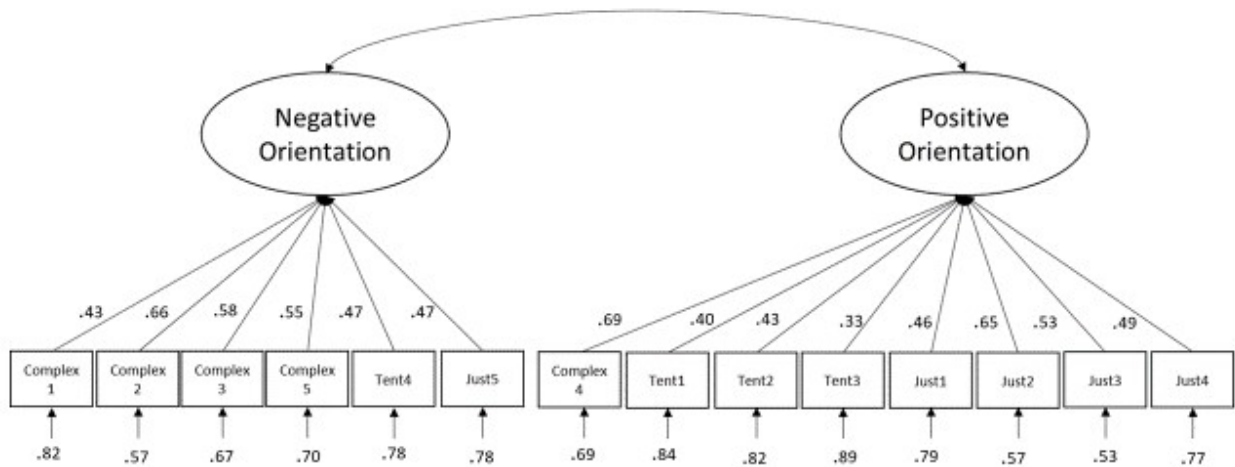
Discussion

Developmental approach

I initially hypothesized that students' would have naïve beliefs about the nature of knowledge in history. However, these initial hypotheses presumed reliable measurement of epistemology and cannot be presently addressed due to issues in both the developmental and dimensional approaches to measuring epistemology. While I suspected the developmental perspective would be less useful due to poor discrimination between categories, fewer items, and a lack of attention to the dimensionality of epistemic beliefs, the measure was more problematic than expected. The current iteration of the scale should be revisited in future studies.

Given the ineffectiveness of the developmental scale, alternative assessments of epistemology from a developmental approach seem more promising than self-report measures. For example, teachers might classify students' epistemologies based on their performance on an applied problem-solving task, such as the Livia Problem (Zavala & Kuhn, 2017). Such a task asks students to explain why two historical accounts of an event disagree and to respond to several questions about the nature of historical knowledge and how it is constructed (e.g. Why do these accounts disagree? How can we resolve their disagreements? Can both accounts be right or only one? How can we come to know what *really* happened?) Such an assessment determines a student's epistemology by observing it in their approach to resolving historical disagreements. Given wide issues with survey responses (Schmitt & Stults, 1985; Nunnally & Bernstein, 1994), this approach may be more promising despite the additional efforts and time needed to administer and code such an assessment.

Making epistemic beliefs scales more discriminant, as may occur in the applied problem-solving approach, appear important. The frequency of evaluativist responses in the present study was at odds with extant research which argues that pervasive approaches to teaching history, such as the transmission of fixed and static knowledge and the predominance of textbook-based instruction, serve as context-level barriers to more sophisticated and discipline-appropriate epistemologies (Bain, 2005, 2006; Goldman et al., 2016; Monte-Sano, 2010; Reisman, 2012; VanSledright, 2002). In more general contexts, students and adults have observed preferences for simple and certain narratives. Such preferences can be explained by the cognitive difficulty of revising previous interpretations, or schema, and includes preferences for coherence instead of accuracy, a preference for single-cause explanations of events, and a preference for explanations that are consistent or confirm prior beliefs (Abendroth & Richter, 2020; Kuhn, 2019; Richter &

Maier, 2017). It is possible that students in the present study had more mature epistemological beliefs than hypothesized, but it is not plausible in the magnitude observed in the data. This does indicate problems with the current measure in accurately measuring beliefs.

Coding thinking and writing might also be a promising approach to accurately categorize an individual's epistemological maturity. In history, epistemology may be manifest in a student response that advances a single cause in an explanation of an (i.e. Rosa Parks alone caused the Montgomery Bus Boycott). In this way, coding students' source-based argument essays or having students engage in think-aloud protocols might be better ways to estimate epistemology.

Dimensional approach

Hypotheses related to the dimensions of students' epistemologies in history also presumed reliable measurement of latent constructs. However, the failure to find a satisfactory model of epistemology did not mean there were no important findings. After documenting issues with item orientation in previous research and the superior fit of the positive and negative factor model, I see at least two major implications for research examining the nature of students' epistemology in history and other domains. First, the latent factors proposed by previous research likely capture student response patterns based on question-wording to some extent. These studies, in so far as they make claims about dimensionality or factor structure, are not robust enough to protect against the influence of the method variance (Bråten, & Strømsø, 2009; Bråten et al., 2008; Bråten, et al., 2011; Strømsø et al., 2008; Wiley et al., 2020).

One objective of the present study was to make disagreements in factor structure more clear (Strømsø & Bråten, 2009; Wiley et al., 2020). This was somewhat achieved as the study brought to light a plausible explanation for apparent disagreements in factor structure that should be further investigated while measuring epistemology in history and other domains—the

capacity for item orientation to confound factor structure. Present analyses provided evidence that the large number of studies using the TESBQ and a dimensional approach to measure epistemology should be re-evaluated given the likely influence of item orientation on factor structure, especially as researchers tie specific student outcomes to specific dimensions of epistemology that may exist as artifacts of method variance.

In short, I cannot *disprove* the concern that factor structure in previous research is influenced by uniform item orientation within scales and polarized item orientation *across* scales. This pattern of item orientation—wholly negatively oriented and wholly positive oriented subscales—seems most susceptible to item orientation confounding factor structure. Therefore, testing for the effects of item orientation should be conducted, both in the formation of and validation of scales proposing a latent construct is best represented by multiple dimensions. When reducing or modifying items, researchers should document their reasons for doing so, including item orientation and correlations with retained items. Empirically investigating the influence of item orientation is vital if researchers are concerned with measuring what they intend to—not artificial method variance.

Notably, Bråten and Strømsø (2010) considered the issue of item orientation briefly in another article about the effects of epistemic beliefs on student reading. They used PCA to reduce items in the TSEBQ and note that items with poor factor loadings are discarded:

> "While it was not possible to differentiate between the items that we excluded and the items that we retained with respect to content, it should be noted that all the items excluded from the justification dimension were negatively worded (e.g., I often feel that I just have to accept that what I read about climate problems can be trusted), suggesting that, at least to our participants, rejection of a negatively worded item did not necessarily mean the same as endorsement of a positively worded item" (p. 643).

Despite this observation, the influence of item orientation is not further investigated, despite methods for doing so. Still, the same pattern of using statistical methods to reduce items

and create uniform item orientation within scales persists in research related to the dimensions of epistemic beliefs (For a most recent example see Yli-Panula et al., 2021).

Given the extant literature documenting how item orientation can confound factor structure, the large number of studies using the original scale, which is not robust to critiques of item orientation, should be re-evaluated in this light (Brown, 2003; Dalal & Carter, 2014; Deemer & Minke, 1999; DiStefano & Motl, 2006; Hughes, 2009; Johanson & Osborn, 2000; Roszkowski & Soven, 2010; Weems et al., 2003). Such studies include those that tie specific student outcomes to specific dimensions of epistemology that may exist as artifacts of method variance. For example, in the same study where Bråten and Strømsø (2010) explicitly discard negatively oriented items, they claim to demonstrate the "unique predictability" of different dimensions on students' multiple-texts comprehension (p. 649). Confounded factor structure due to item orientation plausibly undermines the unique predictability of distinct factors.

As a caveat, researchers find dimensions do positively predict other outcomes that theoretically make sense. This suggests epistemology is being measured to some extent. This is evident in the present study with correlations of uniformly oriented items, even those across hypothesized dimensions. Even so, dimensionality remains an issue across studies previously mentioned.

Second, the quality and usefulness of self-reported survey measures or likert response options should be examined given issues of acquiescence bias, social desirability bias, and the influence of item orientation on student responses (Weems et al. 2002; Hughes, 2009). That students respond differently to the two following questions should give researchers pause as to the usefulness of these items to make claims about psychometrically valid constructs: "What is considered to be certain knowledge about topic today may be considered to be false tomorrow"

and "You can be certain that historical explanations are true."

Many researchers argue against the inclusion of negatively worded items due to the influence on factor structure and means (Cole et al., 2019; Foddy, 1993; Hughes, 2009; Lietz, 2008; Roszkowski & Soven, 2010; Weems et al. 2002). Such an item, "Historical explanations should **not** change in light of new information," (Wiley et al., 2020) is problematic when examined in light of these cautions. Another example, "Problems within climate research do **not** have any clear and **unambiguous** solution," (Bråten, & Strømsø, 2009) highlights what Foddy (1993) and Colston (1999) argue should not be used in scale responses; That problems do **not** have clear solutions is written in the negated positive mode, which is unideal; That problems do **not** have **unambiguous** solutions is in the negated negative mode which is less ideal; To combine both is simply confusing.

Given these issues, future scales should include uniform item orientation *across* scales, especially if they intend to make claims about dimensionality. They should also use clear classification systems, and/or test for the influence of negatively oriented items (Roszkowski & Soven, 2010). Additionally, scales should also clarify what is meant by responses in opposite directions and possibly test versions of scales where individuals choose between two positively and clearly worded options. For example, one can represent the following questions with a sliding scale.

Which do you agree with more? From a single set of facts, you can write….

Only one good historical explanation ⟵⟶ Many good historical explanations

A student who sees just one of the above response options might indicate agreement. On their own, each response seems reasonable. Represented as two polar ends of a construct, whether knowledge in history is simple or complex, the student is forced to choose between two

clear epistemological stances. This would also clarify the names for latent dimensions. Currently, researchers describe one dimension as Certainty (instead of Tentative) and another as Complexity (instead of Simplicity), which results in negative correlations between two constructs that could simply be positively correlated with each other and the reasoning skills associated with a more adaptive epistemology. One additional benefit of using two clear endpoints for the question, students can infer the meaning of words or sentence construction that are unclear from the opposite endpoint. This provides context clues for lower-performing readers and English learners.

Limitations and Conclusion

The present study's sample size (n=256) was sufficient to determine the dimensions of epistemology in the general population, but not large enough to test for differences between important subgroups such as EL students. Given that reading skills and processing may influence scale responses (Dueber et al., 2021; Foddy, 1993; Roszkowski & Soven, 2010), attending to the measurement invariance in subsequent studies will be useful. In preliminary analyses, I conducted CFA models that excluded students in grades 6 and 7 and excluded students with low scores on their essays (scores less than 3 on a 6-point scale). These analyses did not produce any differences in the results.

The developmental scale only had four items, which is much too small to conduct item analyses. As discussed, applied approaches to classifying students' epistemologies appear better than the self-report measures used in the current study, but it is possible that a scale with more items and more testing could improve the usefulness of the approach. Little oversight over the testing situation, the measures we administered by teachers as part of classroom instruction, could have also influenced results for both measures.

While I make claims about the influence of item orientation on factor structure that are supported by an analysis of extant research and some empirical findings, more empirical testing is warranted. Only one item from each respective dimension had polarized item orientation relative to other items. A future study will be conducted in which each hypothesized dimension of epistemology has a combination of both positively and negatively oriented items. With sufficient positively and negatively oriented items for each dimension, the study can advance a multidimensional view of epistemology that is robust to the influence of item orientation. It can also empirically show that factor structures advanced by previous research were largely influenced by method variance. Either result is important for those interested in the accurate and reliable measurement of epistemology.

Study 3: What is the relationship between epistemology and source-based argument writing in history?

Purpose

Given initial findings in Studies 1 and 2, the third study intended to use structural equation modeling to examine the relations between student epistemology and important dimensions of students' source-based argument writing (SBAW) in history. The following research question guided this study: *What is the relationship between epistemology and source-based argument writing in history?*

I originally hypothesized that the dimensions of students' epistemology identified through confirmatory factor analysis (CFA) in Study 2 would be related to students' writing ability, but only weakly related when controlling for other variables such as EL status, sex, and grade level. I also expected epistemology to more strongly predict holistic writing quality for older students as their beliefs should be more strongly developed and, therefore, exert greater

influence on their reading and writing processes in history. Given the issues with item orientation plausibly confounding factor structure described in Study 2, the present study first used CFA to test an alternative model of epistemology before using this model in structural regression models to predict the effect of epistemology on different writing outcomes.

Method

Participants

The same participants and measures from Study 2 were used in Study 3 to measure the relationship between epistemology and holistic writing scores (N=256). These students were those who responded to the epistemology survey and completed the pretest writing assessment as part of the ongoing field trial. Additionally, analytic coding data was used to examine relations between epistemology and certain components of student writing—presentation of ideas, evidence use, and historical thinking. For these analyses, a smaller sample was used. Because certain students were selected from the stratified random sampling in Study 1 and not all students completed the epistemology survey measure described in Study 2, initially there were only 75 students who had both analytic coding data and epistemology data. Therefore, for Study 3, an additional 36 students were sampled and their essays were analytically coded. This was done by first selecting the ten teachers who successfully administered the epistemology survey. Then, I randomly selected an additional six students from each grade whose essays were also analytically coded. The resulting sample included 115 students. This relatively small sample precluded analyses that measure the effects of epistemology on all dimensions of writing simultaneously but the sample is sufficient to detect the effects of epistemology on each dimension of writing independently.

Measures

For the writing outcomes, holistic scores and items in the analytic framework related to the presentation of ideas, evidence use, and historical thinking were used. All writing scores, both holistic and analytic, were created through reliable and rigorous coding (See the measures section in Study 1 for more details). Data used to measure epistemology come from the dimensional epistemology scale described in Study 2.

Analytical approach

This study used multiple structural regression in Mplus 8.4 (Muthén & Muthén, 2017) to predict the effects of epistemology on student writing. Originally, I intended to use the best fitting model of writing in Study 1 and the best fitting model of epistemology in Study 2 as models in the regression. However, due to previous issues with item orientation in Study 2 and the low amount of students for whom there was both epistemology data and analytic coding data, reduced models of epistemology and writing were necessary. Therefore, CFA was conducted before structural regression.

Confirmatory Factor Analysis

The model of SBAW in history described in Study 1 was reduced to include only three factors—Ideas, Evidence Use, and Historical Thinking. Researchers have claimed that epistemology in history is related to how knowledge claims are justified, evidentiary thinking, endorsing competing interpretations, and sourcing moves, but not necessarily related to the structure of an academic essay or a student's academic language (Ashby, Gordon, & Lee, 2005; Bain, 2005; Goldman & Scardamalia, 2013; VanSledright, & Maggioni, 2016). Therefore, items related to Language Use and Structure were removed from analyses. Additionally, analytic coding with high rates of interrater agreement is labor-intensive and costly, so we focused our efforts on those components that were predicted to be relevant by the extant research.

Given the fit of the previous four-factor model of student SBAW in history and the non-significant prediction of the Language Use factor on holistic scores, I expected a three-factor model of SBAW in history to fit the data well. Still, I tested a unidimensional model of writing and compared the hypothesized three-factor model with this baseline model to evaluate the comparative fit of each model.

Next, the model of epistemology was reduced due to the influence of item orientation on factor structure reported in Study 2. While it is evident item orientation influences factor structure to some extent and, therefore, previously hypothesized factors should be re-evaluated (e.g. whether Complex/Tentative is one factor or two), this does not necessarily mean epistemology is *not* being measured by the scale. Indeed, much research confirms the influence of epistemology, albeit with plausibly confounded factor structures, on student reasoning, reading, and writing outcomes (see Greene et al., 2018 for a review). Therefore, I tested an alternative model of epistemology that avoided some issues with item orientation and retained items most likely connected to students' epistemologies in the present writing sample.

First, the items measuring whether knowledge was certain or tentative were removed. This was done because the dissociability of this dimension is least plausible given the results of Study 2. Prior research is unclear as to whether beliefs that knowledge is tentative are separate from beliefs knowledge is complex. This lack of clarity is likely connected to the aforementioned issues with item orientation as Complex and Tentative factors are distinct or a single factor depending on item wording. In the present study, most Tentative items were positively oriented and most Complex items were negatively oriented. Consequently, representing these as factors in CFA may permit the validation of "methods" factors, especially if the one polarized item in each factor was removed due to "poor fit." In other words, to include both factors would be to ignore

the influence of item orientation that was previously discussed. Also, if Tentative and Complex items were set to load onto a single factor in CFA, this same method variance would result in poor model fit.

The removal of the Tentative items was also justified because this dimension is less plausibly connected to student writing in the present writing sample. A belief that knowledge in history is *complex* is essential to writing well in the prompts used presently; students who think there are multiple causes for events are likely to perform well on a task of a causal analysis with multiple sources. Similarly, a belief that knowledge claims in history require integration of evidence from multiple sources, versus using prior experience or authority, is adaptive to writing well in the present SBAW task. Therefore, the Complex and Integration items were retained.

However, a belief that knowledge can change over time is less obviously linked to a single piece of student writing completed in an on-demand testing situation (Bråten & Strømsø, 2006; Christensen-Branum et al., 2019; Strømsø et al., 2008). A task more likely to elicit the bolstering impact of tentative beliefs in knowledge would ask students to revise or reform their interpretation of the past (Kardash & Scholes, 1996). While such a task represents the disciplinary practices of history and requires tentative beliefs in the nature of knowledge, students in the present study did not respond to this type of task.

Next, an item from each of the Complex and Integration subscales was removed. "What is written in textbooks is not necessarily true," was removed because it had poor item fit in the previous factor analyses and analysis of the content of the item indicated it was unclear which response was positively oriented. Sometimes it is reasonable to assume information in textbooks is accurate and true, but at other times, textbook explanations are rewritten to include a revised understanding of the past.

The item, "You need multiple perspectives to understand the causes of historical events," was removed due to poor fit in previous models. Notably, this item was oriented positively, whereas other Complex items were oriented negatively. The remaining respecified model of epistemology included four items for the Complex dimension and four items for the Integration dimension. It is worth noting here and in the discussion that each of these latent factors uses uniform item orientation within scales and polarized item orientation across scales—a problematic pattern shared with previous research that should be considered as a limitation. Still, as discussed previously, it is plausible that epistemology is being measured to some extent. Further, the decision to remove poor fitting items, which results in uniform item orientation within a scale, was frequently made in previous studies (Bråten & Strømsø, 2010; Strømsø & Bråten, 2009; Wiley et al., 2020; Yli-Panula et al., 2021).

Structural Regression

After conducting CFA, four sets of structural regression models were used to show the effect of epistemology on four writing outcomes: holistic scores and each of the three dimensions of writing—Ideas, Evidence Use, and Historical Thinking. Each model was run without controls before a second structural regression model was run that included controls—a dummy variable for age (whether a student was in high school), a dummy variable for sex, and a dummy variable for EL status (designated as EL by the district). Figure 19 shows relations between epistemology and writing outcome with controls added. Such a model allows sex, EL status, and age to directly and indirectly contribute to writing quality through epistemology. The model also shows the direct effect of epistemology on writing controlling for sex, EL status, and being in high school.

Figure 19.
*Relations between Epistemology and Writing Outcomes*
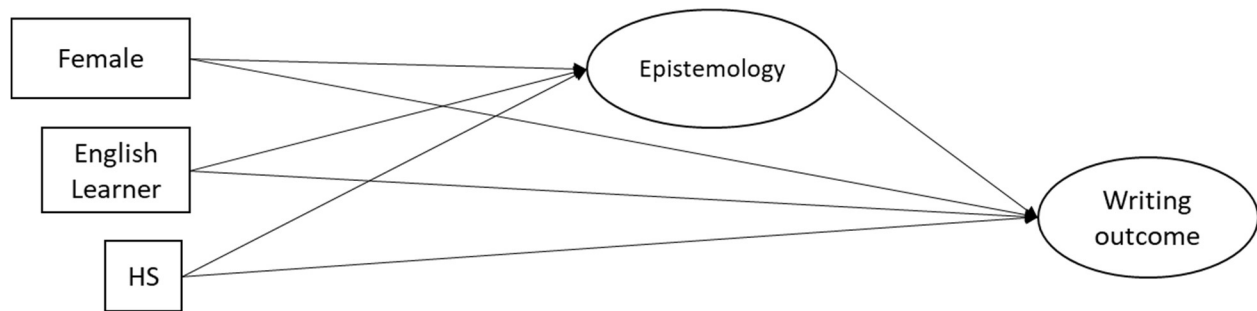
In confirmatory factor analysis and structural regression models, multiple indices were used to evaluate model fit: chi-square statistic, comparative fit index (CFI), root mean square error of approximation (RMSEA), and standardized root mean square residuals (SRMR). RMSEA values below .08, CFI and TLI values equal to or greater than .95, and SRMR equal to or less than .05 indicate excellent model fit (Hu & Bentler, 1999). CFI values greater than .90 and SRMR equal to or less than .10 are considered acceptable (Kline, 2015).

Findings

Confirmatory Factor Analysis

Both the one-factor and three-factor models for SBAW in history had an acceptable fit. However, the three-factor model had superior fit to the one-factor model as indicated by the Chi-square difference of fit test ($\Delta\chi^2$=40.717, df = 4, $p < .001$), superior CFI and TLI values (.946 and .922 versus .920 and .895), superior SRMR value (.045 versus .167), and superior RMSEA value (.133 versus .155). Though RMSEA values were not ideal, the other fit statistics suggest good fit for the model. All factor loadings for items significantly loaded onto their respective factors (.647 $\leq$ .940) and factors were strongly and significantly correlated with each other (.901 $\leq$ .970). See Figure 20.

Figure 20.
*Factors of SBAW in History used in Structural Regression*
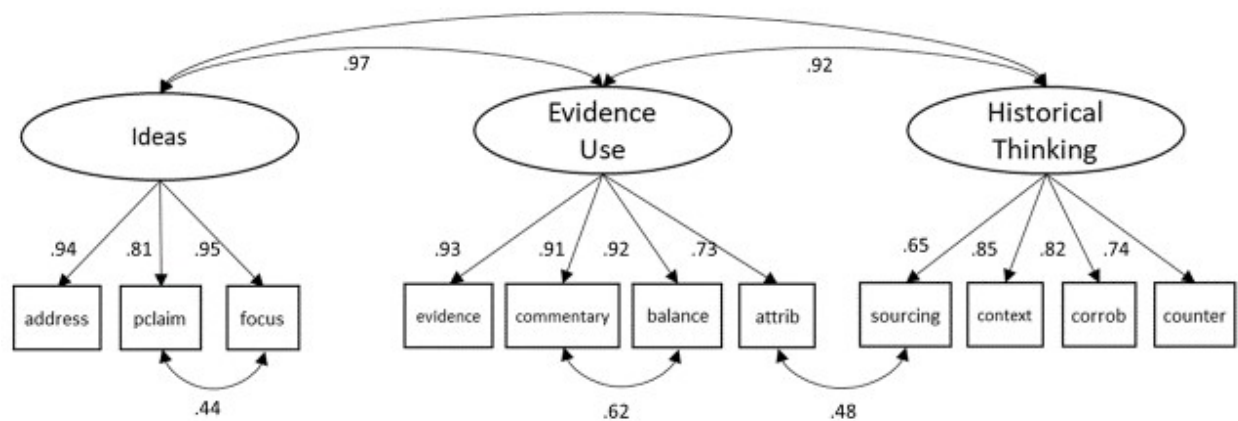
Note. Standardized estimates are reported.

For epistemology, the respecified model had poor fit (χ^2 =60.937, df = 19, p < .001). The

RMSEA value was marginally poor (.093) with values close to 0 and smaller than .08 considered

ideal. CFI and TFI values were also poor (.855 and .787) with values close to 1 and above .9

considered good model fit. The SRMR value (.069) was ideal with a value of < .10 considered

ideal (Hu and Bentler, 1999). This marginally poor fit was still better than many previous models

tested in Study 2. Further, coefficient alpha for the subscales represented in the two-factor CFA

model were .68 and .61, which are similar to coefficients considered acceptable by previous

research (Bråten, & Strømsø, 2009; Wiley et al., 2020). Respecification did not improve model

fit.

Factor loadings for items significantly loaded onto their respective factors (.348 ≤ .726).

Factors were moderately and significantly negatively correlated with each other: -.441.

Theoretically, the negative correlation makes little sense and is likely due to the polarized item

orientations across respective factors and lower mean scores for negatively oriented items—in

this case, items related to the Complex dimension (Weems et al. 2002; Hughes, 2009; Sliter &

Zickar, 2014). Despite the poor fit, given the factor loadings, improved fit compared to models

used in Study 2, and that both dimensions, Complex and Integration, were theoretically related to writing outcomes, this model was selected for use in subsequent structural regression models

Figure 21.
The Preferred Model of Epistemology used in Structural Regression
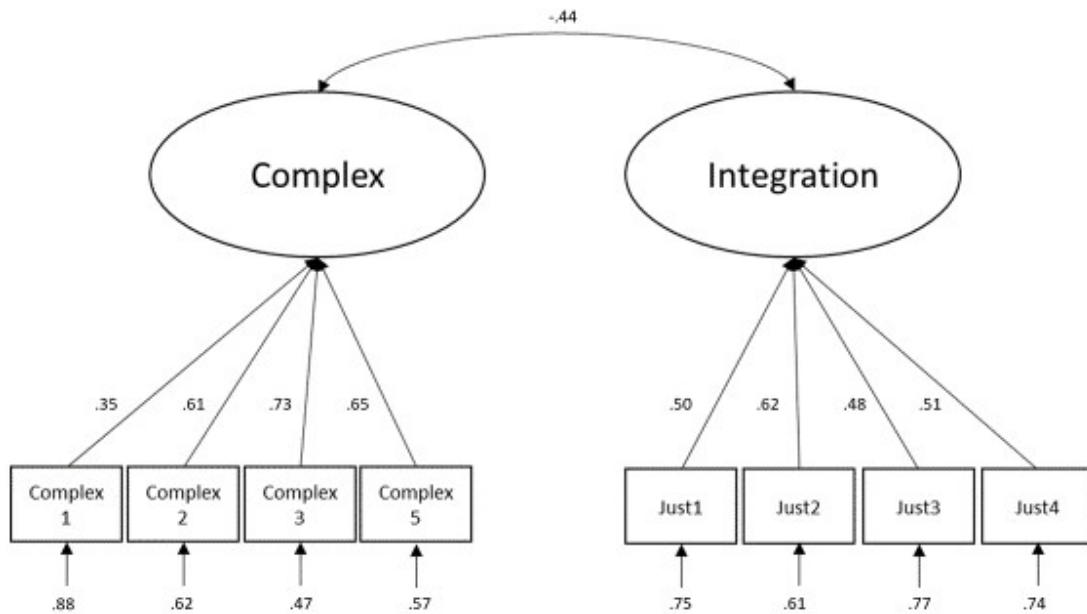
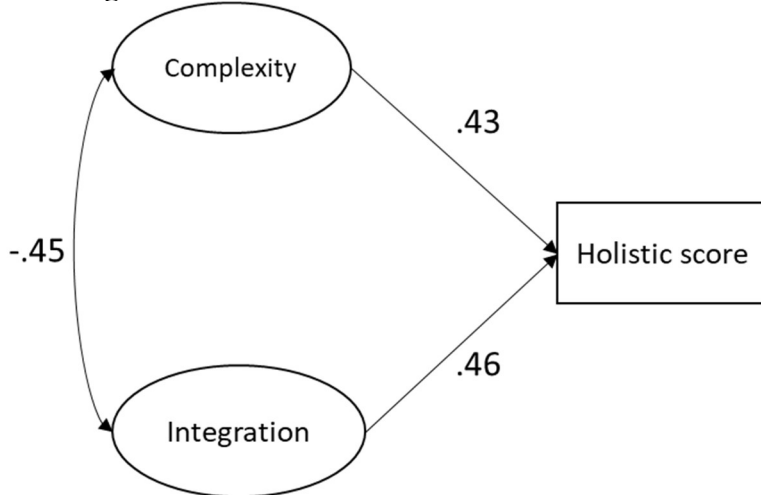*Note.* Standardized estimates are reported.

Structural Regression

To examine the relationship between epistemology and writing quality, four sets of structural regression models were fit to the data. In each set, the writing outcome was first predicted by epistemology before controls were added. In the first set, the preferred model of epistemology was used to predict holistic scores.

Epistemology predicting holistic scores

Figure 22 shows the effects of Complex and Integration factors on students' holistic writing scores. Model fit was good, overall, with individual fit indices ranging from acceptable to good.

Figure 22

Beliefs in the Complexity of Knowledge and Justification through the Integration of Evidence Predicting Students' Holistic Scores

Note. Standardized estimates are reported.

The model shows that both Complex ($b = .427$, $SE = .134$, $p < .01$) and Integration ($b = .464$, $SE = .138$, $p < .01$) factors significantly and positively predicted students' holistic scores. A standard deviation increase on the Complex dimension predicted, on average, a .427 standard deviation increase in a student's holistic score. Similarly, a standard deviation increase on the Integration dimension predicted on average a .464 standard deviation increase in a student's holistic score.

In non-standardized terms, the model predicted that a student scoring a point higher than the average student on the Complex dimension of epistemology (on a scale of 1-10) would score .50 points higher on their writing (on a scale of 1-6). Similarly, a student who scored a point higher on the Integration dimension scored .54 points higher on their writing. Overall, the model explained 22% of the variance in student scores.

With controls added to model—sex, EL status, and being in high school—the significant relations between each dimension of epistemology and the holistic score remained (Complex: $b = .424$, $SE = .176$, $p < .05$ and Integration $b = .430$, $SE = .194$, $p < .05$). The model also showed that students designated EL scored lower on the Integration subscale compared with their non-

EL peers (b = -.389, SE = .110, p < .01), but there was no direct relation between EL status and the holistic score. In the model, students designated ELs did score lower than their peers, but only through lower scores on the Integration factor which in turn predicted writing quality.

The model also showed high school students scored higher on holistic scores compared with middle school students (b = .247, SE = .076, p < .01) and that female students had higher scores on the Complex dimension (b = .227, SE = .019, p < .05), but there were no direct relations from sex to holistic scores or the Integration subscale. Overall, the model with controls explained 26% of the variation in student scores.

Epistemology predicting performance in Evidence Use

The next set of structural regression models used epistemology to predict students' skills in Evidence Use. The models with and without controls had excellent model fit with all fit indices, RMSEA, TLI, CFI, and SRMR, considered ideal. In the first model, both dimensions of epistemology, Complex (b = .523, SE = .144, p < .01) and Integration (b = .610, SE = .146, p < .01), significantly predicted students' performance in Evidence Use. This model explained 35% of the variation in the Evidence Use factor.

When adding controls, as seen in Figure 23, both dimensions of epistemology, Complex (b = .604, SE = .202, p < .01) and Integration (b = .655, SE = .219, p < .01), still significantly predicted students' performance in Evidence Use controlling for sex, being in high school, and EL status. Female students had significantly higher scores on the Complex dimension (b = .227, SE = .109, p < .05), which in turn predicted higher scores in Evidence Use, but there was no direct effect from sex to the Integration factor. Similarly, there was no significant direct effect on Evidence Use for being designated as an EL, but there was a significant negative effect on Integration which in turn predicted students' Evidence Use (b = -.383, SE = .111, p < .01).

116

Figure 23.

Beliefs in the Complexity of Knowledge and Justification through the Integration of Evidence Predicting Students' Evidence Use with Controls
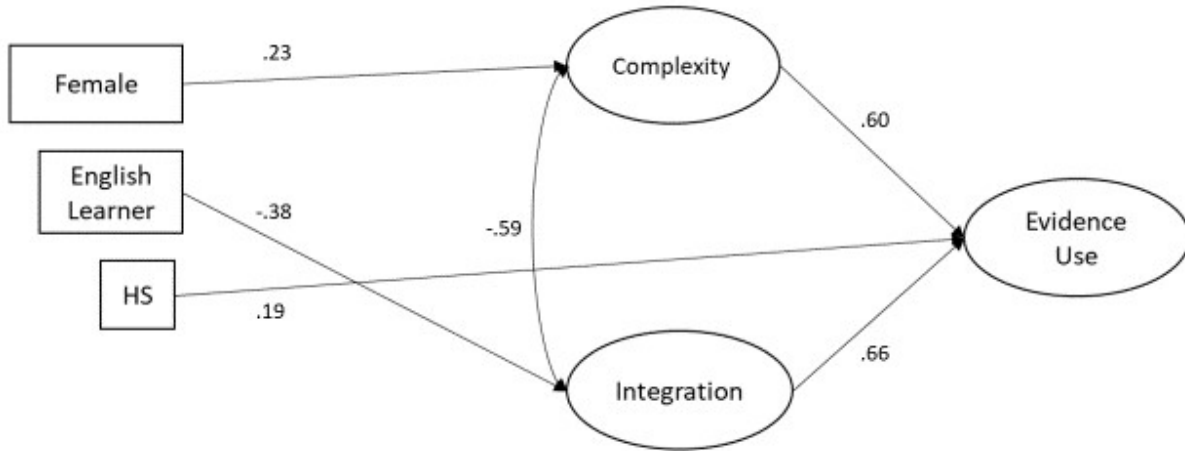
Note. Standardized estimates are reported. Only significant relations are shown in the model.

Lastly, there were no significant relations between being in high school and epistemology, but being in high school did significantly predict a higher score in the Evidence Use factor ($b = .192$, $SE = .092$, $p < .05$). The model with controls added 40% of the variation in the Evidence Use latent factor.

Epistemology predicting performance in Presentation of Ideas

The models showing the relation between epistemology and the Ideas factor had excellent model fit with all fit indices, RMSEA, TLI, CFI, and SRMR, considered ideal. In the first model both dimensions of epistemology, Complex ($b = .737$, $SE = .118$, $p < .01$) and Integration ($b = .705$, $SE = .118$, $p < .01$), significantly predicted students' performance in the Ideas factor. The model with no controls explained 50% of the variation in the Ideas factor.

When adding controls (see Figure 24), both dimensions of epistemology, Complex ($b = .751$, $SE = .150$, $p < .01$) and Integration ($b = .700$, $SE = .167$, $p < .01$), still significantly predicted students' performance in Ideas after controlling for sex, age, and EL status. Female

students had significantly higher scores on the Complex dimension ($b = .222$, $SE = .109$, $p <$ .05), which in turn predicted higher scores in Presentation of Ideas. There was no direct effect from sex to the Ideas factor. Similarly, there was no significant direct effect on Ideas for being designated as an EL, but there was a significant negative effect on Integration which, in turn, predicted students' scores in the Ideas factor ($b = -.392$, $SE = .111$, $p < .01$) Unlike the previous three models, there were no significant relations between being in high school and any dependent variables in this model, including the Ideas factor.

Figure 24.
Beliefs in the Complexity of Knowledge and Justification through the Integration of Evidence Predicting Students' Presentation of Ideas with Controls
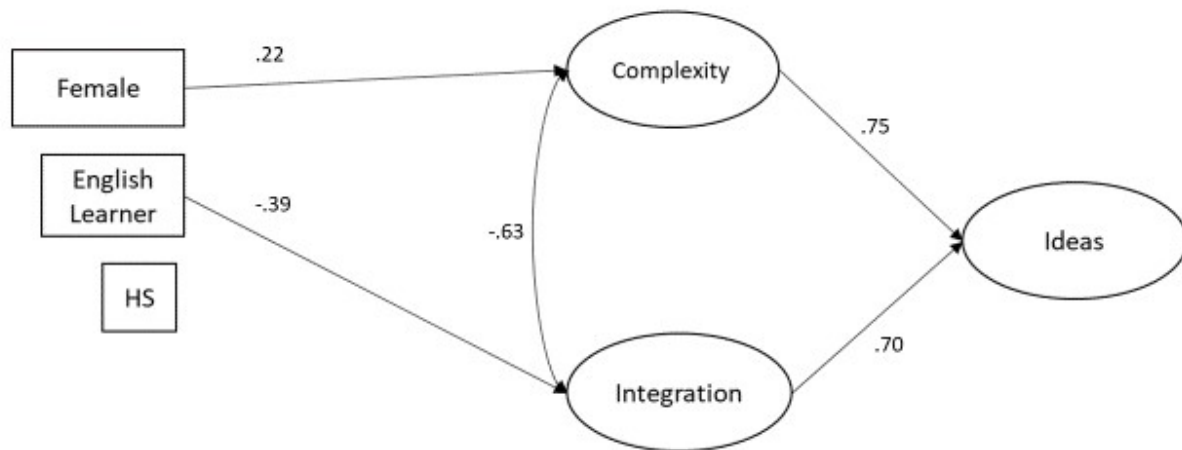
Note. Standardized estimates are reported. Only significant relations are shown in the model.

Overall, the model with controls predicted 45% of the variation in the Presentation of Ideas factor.

Epistemology predicting performance in Historical Thinking

The final set of models showed the effects of epistemology on students' Historical Thinking. The models with and without controls both had good model fit with all fit indices, RMSEA, SRMR, TLI, and CFI considered ideal. In the first model, both dimensions of

epistemology, Complex (b = .368, SE = .140, p < .01) and Integration (b = .371, SE = .148, p <

.05), significantly predicted students' performance in Historical Thinking. The model without

controls explained 15% of the variation in the Historical Thinking factor.

When adding controls, neither dimension of epistemology significantly predicted scores

on the Historical Thinking factor: Complex (b = .303, SE = .188, p = .107) and Integration (b =

.271, SE = .211, p = .198). The model also showed that female students had significantly higher

scores on the Complex dimension (b = .225, SE = .110, p < .05), but there was no significant

effect from sex to Historical Thinking. Similarly, there was no significant direct effect on

Historical Thinking for being designated as an English Learner, but there was a significant

negative effect on Integration (b = -.402, SE = .110, p < .01). There were no significant relations

between being in high school and epistemology, but being in high school did significantly

predict a higher score in the Historical Thinking factor (b = .168, SE = .075, p < .05). The model

explained 16% of the variation in Historical Thinking.

Discussion

Students who believed that knowledge in history is complex, not simple, and valued

integration when forming historical explanations reported increased writing achievement on a

source-based argument writing prompt in history class. This finding is in line with previous

studies and the use of multiple continuous indicators of writing quality permits for more accurate

estimates of the effect of epistemology on writing quality (Bråten et al., 2011; Greene et al.,

2018).

The writing task in which students engaged, required thinking about multiple potential

forces, actors, and events that led to the success of an organized social/political movement. To

respond to the prompt well, students had to think about myriad actors and events that interacted

over time; they had to reject a simple and clear explanation. Therefore, a belief that history is complex, that multiple, competing and complementary forces produce historical outcomes, produced improved historical writing as seen presently. Similarly, because the SBAW task required a writer to evaluate multiple sources, and these sources represent various actors and events leading to the outcome of a social/political movement, a belief in the value of integrating evidence to justify claims also leads to improved writing. Multiple sources needed to be used to answer the prompt accurately and to support whatever claim students made with sufficient evidence. For example, consider the following excerpt from a student:

> Even though people may argue that important people like Cesar Chavez and Dolores Huerta were the overall cause for the success of the strikes, the great efforts of all of the grape workers involved truly made the success of the Delano Strike and Boycott possible. Cesar Chavez and Dolores Huerta are famously recognized in leading the strikes and boycotts. But without the united and consistent efforts of all the people who participated, they would have never seen success.

While the writer could have made the simple claim that Chavez's advocacy is the reason the boycott succeeded, in endorsing a more complex explanation the writer acknowledges historical complexity inherent to the movement's progress and is positioned to write an effective source-based argument essay. The writer shows an understanding that historical forces (e.g. Huerta, Chavez, all grape workers) vary in their influence and interact with other forces in their historical contexts. A complex view of historical knowledge is manifest in a clear and compelling claim that is aligned with multiple pieces of evidence in the sources.

The writing also references another perspective (that Chavez and Huerta were the most important factors leading to success) that can be addressed in the body of the essay for a compelling historical argument. This excerpt shows the connection between epistemology and writing, in this case, the Ideas factor. Those interested in supporting strong historical writing might find it helpful to develop students' views of knowledge as complex. Once students

understand that history knowledge is complex, they will think—and therefore write—differently. One way to move students towards this understanding of the complexity of history would be to use such a claim like the one quoted as a model and explicitly instruct students to account for multiple forces or causes in their claims.

Another excerpt from student writing illustrates the connections between the Integration factor and student writing:

> Someone might argue that people coming together to join in the grape boycott was the most important reason for the success of the strike and boycott. I would defend my argument by referencing the headnote in source 3 which states… "César Chavez speaks on TV and writes letters asking people to stop buying all California grape Communicating in many ways was important. If people just happened not to get the letters, at least they were able to be informed by the TV. César Chávez found a way for everyone to know about the grape strike and the boycott. Spreading the message by TV allowed many people to see and then act.

This student is leveraging skills with evidence use, presenting their ideas, and attending to the historical context to support their argument. These skills are directly supported by a belief in the value of integrating evidence from multiple sources to support a claim, as indicated by this student's reference to "source 3." While many students in the sample referenced a potential counterargument to their own claims, not all students refuted this counterargument with evidence and further analyses. The excerpt above shows a student who refutes a counterargument with evidence and analysis that is indicative of high-quality SBAW. Therefore, a belief in integrating evidence to support and defend claims can produce better writing, especially as it facilitates practice and uptake of evidentiary thinking and argumentation.

Turning to the models with controls, these models showed potential moderating effects between epistemology and the writing performance of certain students, though models with more robust controls are necessary to make such claims with conviction. Females consistently scored higher on the Complex dimension and this, in turn, predicted holistic scores and performance on

the Evidence Use and Ideas factors. This suggests complex beliefs about the nature of knowledge in history could explain, at least partially, the higher writing scores of female students.

There also appeared to be no differences between the epistemologies of HS and MS students. High school students' higher holistic writing scores and outperformance in Evidence Use and Historical Thinking factors occurred directly, not through epistemology as a mediator. Another notable finding includes the consistently lower scores for students designated EL on the Integration factor, which in turn predicted lower writing performance. Coupled with the lack of significant direct effects from EL Status on writing scores, this suggests responses on the Integration subscale, in part, explain the relation between EL Status and writing.

That epistemology did not significantly predict Historical Thinking once controls were added to the model is notable. The magnitude of the relations in the model without controls was also lower for Historical Thinking than for other factors. One explanation is the lower observed scores for items related to Historical Thinking in the present sample. It is plausible that skills like sourcing, contextualization, and presenting counterarguments are linked to epistemology even after controlling for other factors, but there skills are not developed enough in the present study to see such an effect. Study 1 showed that students struggled to display these historical writing skills.

Finding also have important implications for teachers. When engaging in historical inquiry, it is important to explicitly teach that there will be multiple, competing interpretations to inquiry questions. Giving time for students to identify different responses to questions about the past supports the development of beliefs that knowledge is complex (De La Paz et al., 2017; Monte-Sano, 2010; Nokes, 2013; Young & Leinhardt, 1998). Then, allowing students to make sense of different sources of evidence and prompting students to use this evidence to evaluate the

strength of different interpretations should directly help students make stronger written arguments later (Wiley et al., 2020). Helping students think about history as complex historical knowledge as supported by integrating multiple sources of evidence appears to be a useful endeavor for teachers.

Limitations and Conclusion

One major limitation of this study is the poor fitting model of epistemology and the use of factors that are plausibly confounded with item orientation. At the same time, the model of epistemology used to predict writing is not substantively different from those used in prior research. The principal difference is that in the present study issues with item orientation were explicitly addressed instead of ignored.

Still, if negatively oriented items are causing distinct response patterns, this affects claims about factor structure and construct validity (Deuber et al., 2021, p.2). Future studies should try to account for this by redesigning measures to account for patterns in response bias or using methodological approaches to account for the variance attributed to item orientation. Such an approach, like the use of bifactor models with specific factors representing item orientation, requires a larger sample size but should be tested in subsequent studies (DiStefano & Motl, 2006).

The small sample size, especially for analytic coding data, is another issue in the present study. While such coding is time-consuming, the varying effects of epistemology observed in the Ideas, Evidence Use, and Historical Thinking factors indicate the value of such an approach. Study 1 also shows how skills vary among subgroups of students. With a larger sample of students and more analytic coding, a more robust model of the effects of epistemology on student writing could be pursued. Additionally, a sample with more variation in writing scores might

also better detect a relation between Historical Thinking and epistemology. Finally, regression models with more robust covariates, such as controlling for reading comprehension, would make claims about the influence of epistemology more robust.

Overall, despite methodical limitations, there does appear to be a relationship between students' beliefs about knowledge and their writing. Such a relationship should be acknowledged in future studies that try to improve student writing. Because a view of knowledge as complex and validated by evidentiary thinking is foundational to making arguments about the past, teachers are encouraged to incorporate more source-based inquiry tasks in their classrooms, especially those with multiple, complex, and unclear answers. Because writing arguments about the past using multiple sources is complex and challenging, teachers also need instructional support, such as high-quality source-based inquiry units and professional development that helps teachers develop students' reading, thinking, and writing skills.

Contributions to the Field

In this dissertation, I used a rigorous analytic human coding scheme, developed and applied by a team of researchers, to offer a detailed and analytic description of students' source-based argument writing in history. An examination of how students performed across discrete skills, like sourcing and structuring the body of an essay, and how performance varied by subgroups of students had not been yet observed in extant literature (to my knowledge). Further, while previous research had examined students' overall writing quality and their historical thinking as separate constructs, CFA offered a picture of them as separate but strongly related. These findings have implications for both writing assessment and practice.

A multidimensional view of SBAW in history validates the use of analytic rubrics and measurement of discrete skills, especially as differences in these skills are observed across grade

levels (i.e. 7th-grade students can state a claim, but cannot yet articulate a counterargument). For instructional purposes, the descriptive statistics of student writing across items, paired with the factor structure, indicate what skills should be targeted by instruction. Findings suggest older students should work on developing reasoning and historical thinking skills that appear to plateau in their development in the present sample. Historical Thinking and Evidence Use contribute significantly to high school students' overall writing quality as measured by a holistic score.

The findings of a relationship between EL status and the factors of SBAW in history also contribute to a growing body of research that shows ELs need a comprehensive approach to writing instruction that focuses on the development of ideas and interpretative, evidentiary thinking rather than one solely focused on language skills (Olson et al., 2020; Fitzgerald, 2017; Short & Fitzsimmons, 2007). Further, instructional approaches that are only language-based take a deficit view of these learners and may exacerbate opportunity gaps (Applebee et al., 2003).

Study 2 made a significant contribution to a growing field studying how different dimensions of epistemology predict student reading, reasoning, and writing outcomes. Findings suggest that models of epistemology advanced by previous research have factor structures that are plausibly confounded by item orientation. While a more comprehensive review of research using extant scales is needed, it also appears little has been done to guard against the influence of method variance when constructing dimensional scales to measure epistemology. Further, this method variance likely contributed to prior findings that asserted the unique predictability of distinct dimensions of epistemology. Therefore, serious and systematic changes need to be applied to scales measuring dimensions of epistemology in history and other domains. A follow-up study has already been planned that features a sufficient number of positively and negatively

oriented items, drawn from previous research, to confirm the emerging hypothesis that item orientation confounds the factor structure of epistemology in prior research

Study 3, though limited by measurement issues raised in Study 2, offers evidence that epistemology is related to key SBAW skills in history. A belief that knowledge in history is complex was related to students' holistic scores and specific factors of SBAW—Ideas, Evidence Use, and Historical Thinking. Similar relations existed for beliefs related to the integration of evidence to support claims. As SBAW in history means resolving multiple competing interpretations of the past through evidentiary reasoning, a view of knowledge as complex and justified through multiple sources of evidence is adaptive for writing well in this discipline. Educators and researchers concerned with improving SBAW in history are, therefore, justified in targeting students' beliefs about knowledge in history. That epistemology was related to multiple factors of students' SBAW also speaks to the interrelatedness of skills like selecting and analyzing evidence, contextualization, sourcing, and presenting clear claims. This is just one more area where we see that writing is, indeed, a complex skill (Hayes, 2012).

Future studies should examine the malleability of epistemology as well as the malleability of key writing skills. For example, a study examining if writing improvement is moderated by epistemology will offer further insight into the importance of epistemology in SBAW development. Future studies will also benefit from improved measurement of epistemology, both in controlling for biased response patterns in self-report measures and including other dimensions of epistemology such as a belief that knowledge is tentative. A belief that knowledge is tentative, that it can change over time, is likely linked to integrating new information about past events or revising one's understanding of historical narratives. Writing tasks that ask students to *revise* their understandings of the past are plausibly influenced by

epistemology and should also be investigated.

Improving students' thinking about history can help them become better writers. Specifically, teachers can help students think about history as complex, competing interpretations of the past that are resolved by integrating multiple, reliable sources of information. In doing so, teachers are also preparing students to reason with multiple conflicting sources of evidence, evaluate the reliability of sources of information, and consider multiple explanations or claims before advancing an argument. In this way, helping students think well and argue well is also linked to students' present and future civic reasoning and discourse (McGrew et al., 2018). For example, developing students' beliefs that claims should be justified with multiple sources of evidence and developing students' skills in "sourcing" are vital to combat misinformation that threatens to erode democratic society.

Researchers have argued for an emphasis on epistemic beliefs in civic contexts (Barzilai and Chinn, 2020). For example, Chinn and colleagues (2020) argue that to improve civic discourse, educators should explicitly consider the following with students: "what are the norms and beliefs we should endorse to effectively answer historical, civic, and social questions?" Students' beliefs about knowledge influence their reasoning in historical and civic contexts whether educators acknowledge this or not. Therefore, a model of student development that acknowledges the role of epistemology is needed to fully understand student reading, thinking, and writing.

As researchers and educators move forward in helping students develop the source-based arugment writing skills needed for 21st-century society, it is vital to acknowledge 1) the connection between beliefs and source-based arugment writing; 2) the connection between thinking historically and thinking civically, and 3) the role that history teachers have in building

these competencies, beliefs, and skills in students. With this more complete model of developing

students' "argument literacy" (Graff, 2003), we prioritize students' futures as careful and critical

thinkers in college, the workplace, and civic life.

References

Abendroth, J., & Richter, T. (2020). Mere plausibility enhances comprehension: The role of plausibility in comprehending an unfamiliar scientific debate. *Journal of Educational Psychology*.

Alexander, P. A., Dinsmore, D. L., Fox, E., Grossnickle, E. M., Loughlin, S. M., Maggioni, L., Parkinson, M., & Winters, F. I. (2011). Higher order thinking and knowledge: Domain-general and domain-specific trends and future directions.

Anastasi, A., & Urbina, S. (1997). *Psychological testing*. Prentice Hall/Pearson Education.

Anmarkrud, Ø., Bråten, I., & Strømsø, H. I. (2014). Multiple-documents literacy: Strategic processing, source awareness, and argumentation when reading multiple conflicting documents. *Learning and Individual Differences*, *30*, 64-76.

Applebee, A. N., & Langer, J. A. (2011). " EJ" Extra: A Snapshot of Writing Instruction in Middle Schools and High Schools. *The English Journal*, *100*(6), 14-2+7.

Applebee, A. N., Langer, J. A., Nystrand, M., & Gamoran, A. (2003). Discussion-based approaches to developing understanding: Classroom instruction and student performance in middle and high school English. *American Educational research journal, 40*(3), 685-730.

Ashby, R., Gordon, P., & Lee, P. (2005). Understanding history. *London: RoutledgeFalmer*.

Bacha, N. (2001). Writing evaluation: what can analytic versus holistic essay scoring tell us?. *System, 29*(3), 371-383.

Bain, R. B. (2005). They thought the world was flat: Applying the principles of how people learn in teaching high school history. *How students learn: History in the classroom*, 179-213.

Bain, R. (2006). Rounding up unusual suspects: Facing the authority hidden in the history

    classroom. *Teachers College Record*, *108*(10), 2080-2114.

Bandalos, D. L. (2018). *Measurement theory and applications for the social sciences*. Guilford

    Publications.

Bang, H. J. (2013). Reliability of national writing project's analytic writing continuum

    assessment system. *Journal of Writing Assessment*, *6*(1), 13-24.

Barnette, J. J. (2000). Effects of stem and Likert response option reversals on survey internal

    consistency: If you feel the need, there is a better alternative to using those negatively

    worded stems. *Educational and psychological measurement*, *60*(3), 361-370.

Barzilai, S., Tzadok, E., & Eshet-Alkalai, Y. (2015). Sourcing while reading divergent expert

    accounts: Pathways from views of knowing to written argumentation. *Instructional*

    Science, *43*(6), 737-766.

Barzilai, S., & Eshet-Alkalai, Y. (2015). The role of epistemic perspectives in comprehension of

    multiple author viewpoints. *Learning and Instruction*, *36*, 86-103.

Barzilai, S., & Chinn, C. A. (2020). A review of educational responses to the "post-truth"

    condition: Four lenses on "post-truth" problems. *Educational Psychologist*, *55*(3), 107-

    119.

Barzilai, S., & Zohar, A. (2012). Epistemic thinking in action: Evaluating and integrating online

    sources. *Cognition and Instruction*, *30*(1), 39-85.

Barzilai, S., & Zohar, A. (2014). Reconsidering personal epistemology as metacognition: A

    multifaceted approach to the analysis of epistemic thinking. *Educational psychologist*,

    49(1), 13-35.

Barzilai, S., Zohar, A. R., & Mor-Hagani, S. (2018). Promoting integration of multiple texts: A review of instructional approaches and practices. *Educational psychology review*, *30*(3), 973-999.

Barzilai, S., & Weinstock, M. (2015). Measuring epistemic thinking within and across topics: A scenario-based approach. *Contemporary Educational Psychology*, *42*, 141-158.

Bickford, J. H. (2010). Uncomplicated technologies and erstwhile aids: How PowerPoint, the Internet, and political cartoons can elicit engagement and challenge thinking in new ways. *The History Teacher*, *44*(1), 51-66.

Braasch, J. L., Lawless, K. A., Goldman, S. R., Manning, F. H., Gomez, K. W., & MacLeod, S. M. (2009). Evaluating search results: An empirical analysis of middle school students' use of source attributes to select useful sources. *Journal of Educational Computing Research*, *41*(1), 63-82.

Braasch, J. L., Bråten, I., Strømsø, H. I., & Anmarkrud, Ø. (2014). Incremental theories of intelligence predict multiple document comprehension. *Learning and Individual Differences*, *31*, 11-20.

Braasch, J. L., & Bråten, I. (2017). The discrepancy-induced source comprehension (D-ISC) model: Basic assumptions and preliminary evidence. *Educational Psychologist*, *52*(3), 167-181.

Bråten, I., & Strømsø, H. I. (2006). Effects of personal epistemology on the understanding of multiple texts. *Reading Psychology*, *27*(5), 457-484.

Bråten, I., & Strømsø, H. I. (2010). When law students read multiple documents about global warming: Examining the role of topic-specific beliefs about the nature of knowledge and knowing. *Instructional Science*, *38*(6), 635-657.

Bråten, I., Strømsø, H. I., & Samuelstuen, M. S. (2008). Are sophisticated students always better? The role of topic-specific personal epistemology in the understanding of multiple expository texts. *Contemporary Educational Psychology*, *33*(4), 814-840.

Bråten, I., & Strømsø, H. I. (2009). Effects of task instruction and personal epistemology on the understanding of multiple texts about climate change. *Discourse Processes, 47*(1), 1-31.

Bråten, I., Britt, M. A., Strømsø, H. I., & Rouet, J. F. (2011). The role of epistemic beliefs in the comprehension of multiple expository texts: Toward an integrated model. *Educational Psychologist*, *46*(1), 48-70.

Bråten, I., Strømsø, H. I., & Salmerón, L. (2011). Trust and mistrust when students read multiple information sources about climate change. *Learning and Instruction*, *21*(2), 180-192.

Bråten, I., Ferguson, L. E., Anmarkrud, Ø., & Strømsø, H. I. (2013). Prediction of learning and comprehension when adolescents read multiple texts: The roles of word-level processing, strategic approach, and reading motivation.

Bråten, I., Ferguson, L. E., Strømsø, H. I., & Anmarkrud, Ø. (2014). Students working with multiple conflicting documents on a scientific issue: Relations between epistemic cognition while reading and sourcing and argumentation in essays. *British Journal of Educational Psychology*, *84*(1), 58-85.

Bråten, I., Stadtler, M., & Salmerón, L. (2017). The role of sourcing in discourse comprehension. In *The Routledge handbook of discourse processes* (pp. 141-166). Routledge.

Britt, M. A., Perfetti, C. A., Sandak, R., & Rouet, J. F. (1999). Content integration and source separation in learning from multiple texts. *Narrative comprehension, causality, and coherence: Essays in honor of Tom Trabasso*, 209-233.

Britt, M. A., & Aglinskas, C. (2002). Improving students' ability to identify and use source information. *Cognition and instruction*, *20*(4), 485-522.

Britt, M. A., Rouet, J. F., & Braasch, J. L. (2012). Documents as entities: Extending the situation model theory of comprehension. In *Reading-from words to multiple texts* (pp. 174-193). Routledge.

Breakstone, J., Smith, M., & Wineburg, S. (2013). Beyond the bubble in history/social studies assessments. *Phi Delta Kappan, 94*(5), 53–57. doi:10.1177/003172171309400512

Bromme, R. (2005). Thinking and knowing about knowledge. In *Activity and sign* (pp. 191-201). Springer, Boston, MA.

Bromme, R., Pieschl, S., & Stahl, E. (2010). Epistemological beliefs are standards for adaptive learning: a functional theory about epistemological beliefs and metacognition. *Metacognition and learning*, *5*(1), 7-26.

Brown, T. A. (2003). Confirmatory factor analysis of the Penn State Worry Questionnaire: Multiple factors or method effects?. *Behaviour research and therapy, 41*(12), 1411-1426.

Buehl, M. M., & Alexander, P. A. (2005). Motivation and performance differences in students' domain-specific epistemological belief profiles. *American Educational Research Journal*, *42*(4), 697-726.

Charney, D. (1984). The validity of using holistic scoring to evaluate writing: A critical overview. *Research in the Teaching of English*, 65-81.

Chinn, C. A., & Buckland, L. A. (2012). Model-based instruction: Fostering change in evolutionary conceptions and in epistemic practices. *Evolution challenges: Integrating research and practice in teaching and learning about evolution*, 211-232.

Chinn, C. A., Barzilai, S., & Duncan, R. G. (2020). Disagreeing about how to know: The instructional value of explorations into knowing. *Educational Psychologist, 55*(3), 167-180.

Christensen-Branum, L., Strong, A., & Jones, C. D. O. (2019). Mitigating Myside Bias in Argumentation. *Journal of Adolescent & Adult Literacy*, *62*(4), 435-445.

Cribb, G., Maglio, C., & Greenleaf, C. (2018). Collaborative argumentation: Tenth graders read modern Iranian history. *The History Teacher*, *51*(3), 477-526.

Crossley, S. (2020). Linguistic features in writing quality and development: An overview. *Journal of Writing Research, 11*(3).

Cole, K. L., Turner, R. C., & Gitchel, W. D. (2019). A study of polytomous IRT methods and item wording directionality effects on perceived stress items. *Personality and Individual Differences*, *147*, 63-72.

Common Core State Standards Initiative (2012). *Implementing the common core state standards.* Retrieved from Common Core State Standards Initiative: Preparing America's Students for College and Career: http://www.corestandards.org

Colston, H. L. (1999). "Not good" is "bad," but "not bad" is not "good": An analysis of three accounts of negation asymmetry. *Discourse Processes*, *28*(3), 237-256.

Correnti, R., Matsumura, L. C., Hamilton, L., & Wang, E. (2013). Assessing students' skills at writing analytically in response to texts. *The Elementary School Journal, 114*(2), 142-177.

Correnti, R., Matsumura, L. C., Wang, E., Litman, D., Rahimi, Z., & Kisa, Z. (2020). Automated scoring of students' use of text evidence in writing. *Reading Research Quarterly, 55*(3), 493-520.

Council of Writing Program Administrators, National Council of Teachers of English, & National Writing Project. (2011). Framework for success in postsecondary writing. Retrieved from wpacouncil.org/framework

Crossley, S., & McNamara, D. (2010). Cohesion, coherence, and expert evaluations of writing proficiency. In *Proceedings of the Annual Meeting of the Cognitive Science Society* (Vol. 32, No. 32).

Cowgill II, D. A., & Waring, S. M. (2017). Historical Thinking: Analyzing Student and Teacher Ability to Analyze Sources. *Journal of Social Studies Education Research*, *8*(1), 115-145.

CWPA, N. (2011). NWP (Council of Writing Program Administrators, the National Council of Teachers of English, and the National Writing Project). 2011. *Framework for Success in Postsecondary Writing.*

Dalal, D. K., & Carter, N. T. (2014). *Negatively worded items negatively impact survey research.* In More statistical and methodological myths and urban legends (pp. 112-132). Routledge.

Deemer, S. A., & Minke, K. M. (1999). An investigation of the factor structure of the teacher efficacy scale. *The journal of educational research, 93*(1), 3-10.

De La Paz, S., Ferretti, R., Wissinger, D., Yee, L., & MacArthur, C. (2012). Adolescents' disciplinary Evidence Use, argumentative strategies, and organizational structure in writing about historical controversies. *Written Communication*, *29*(4), 412-454.

De La Paz, S., Monte-Sano, C., Felton, M., Croninger, R., Jackson, C., & Piantedosi, K. W. (2017). A historical writing apprenticeship for adolescents: Integrating disciplinary learning with cognitive strategies. *Reading Research Quarterly*, *52*(1), 31-52.

DiStefano, C., & Motl, R. W. (2006). Further investigating method effects associated with negatively worded items on self-report surveys. *Structural Equation Modeling, 13*(3), 440-464.

Du, H., & List, A. (2020). Researching and writing based on multiple texts. *Learning and Instruction*, *66*, 101297.

Dueber, D. M., Toland, M. D., Lingat, J. E., Love, A. M., Qiu, C., Wu, R., & Brown, A. V. (2021). To reverse item orientation or not to reverse item orientation, that is the question. *Assessment*, 10731911211017635.

Ferguson, L. E., & Bråten, I. (2013). Student profiles of knowledge and epistemic beliefs: Changes and relations to multiple-text comprehension. *Learning and Instruction*, *25*, 49-61.

Flower, L., & Hayes, J. R. (1981). A cognitive process theory of writing. *College composition and communication*, *32*(4), 365-387.

Foddy, W. (1993) *Constructing questions for interviews and questionnaires.* Theory and practice in social research. Cambridge, UK: Cambridge University Press

Gallagher, H. A., Arshan, N., & Woodworth, K. (2017). Impact of the National Writing Project's College-Ready Writers Program in high-need rural districts. *Journal of Research on Educational Effectiveness, 10*(3), 570-595.

Gansle, K. A., Noell, G. H., VanDerHeyden, A. M., Slider, N. J., Hoffpauir, L. D., Whitmarsh, E.L., et al. (2004). An examination of the criterion validity and sensitivity to brief intervention of alternate curriculum-based measures of writing skill. *Psychology in the Schools, 41*, 291–300.

Gibbons, R. D., & Hedeker, D. R. (1992). Full-information item bi-factor analysis.
Psychometrika, 57(3), 423-436.

Graham, S., Harris, K. R., & Mason, L. (2005). Improving the writing performance, knowledge,
and self-efficacy of struggling young writers: The effects of self-regulated strategy
development. *Contemporary educational psychology*, *30*(2), 207-241.

Graham, S., Harris, K. R., & Santangelo, T. (2015). based writing practices and the common
core: Meta-analysis and meta-synthesis. *The Elementary School Journal*, *115*(4), 498-
522.

Graham, S., Harris, K. R., & Chambers, A. B. (2016). Evidence-based practice and writing
instruction. *Handbook of writing research*, *2*, 211-226.

Greene, J. A., Sandoval, W. A., & Bråten, I. (2016). An introduction to epistemic cognition.
Handbook of epistemic cognition, 1-15.

Greene, J. A., Cartiff, B. M., & Duke, R. F. (2018). A meta-analytic review of the relationship
between epistemic cognition and academic achievement. *Journal of Educational
Psychology*, *110*(8), 1084.

Greeno, J.G. (2011). A situative perspective on cognition and learning in interaction. In:
Koschmann T. (eds) Theories of Learning and Studies of Instructional Practice.
Explorations in the Learning Sciences, Instructional

Goldman, S. R., Braasch, J. L., Wiley, J., Graesser, A. C., & Brodowinska, K. (2012).
Comprehending and learning from Internet sources: Processing patterns of better and
poorer learners. *Reading Research Quarterly, 47*(4), 356-381.

Goldman, S. R., & Scardamalia, M. (2013). Managing, understanding, applying, and creating knowledge in the information age: Next-generation challenges and opportunities. *Cognition and Instruction*, *31*(2), 255-269.

Goldman, S. R., Britt, M. A., Brown, W., Cribb, G., George, M., Greenleaf, C., Lee, C., Shanahan, C., & Project READI. (2016) Disciplinary Literacies and Learning to Read for Understanding: A Conceptual Framework for Disciplinary Literacy. *Educational Psychologist, 51(2)*, 219-246.

Gottlieb, E., & Wineburg, S. (2012). Between veritas and communitas: Epistemic switching in the reading of academic and sacred history. *Journal of the Learning Sciences*, *21*(1), 84-129.

Hayes J. R, (2000). *A new framewor*k for understanding a cognition and affect in writing. Perspectives on writing: Research, theory, and practice. Systems and Performance Technologies, vol 1 (pp. 41-71). Springer, New York, NY.

Hayes, J. R. (2012). Modeling and remodeling writing. *Written communication*, *29*(3), 369-388.

Hastings, P., Hughes, S., Magliano, J. P., Goldman, S. R., & Lawless, K. (2012). Assessing the use of multiple sources in student essays. *Behavior Research Methods*, *44*(3), 622-633.

Hillocks, G. (1995). *Teaching writing as reflective practice*. Teachers College Press.

Hillocks, G. (2011). Teaching argument writing, grades 6–12. *Portsmouth, NH: Heinemann*.

Hofer, B. K., & Pintrich, P. R. (1997). The development of epistemological theories: Beliefs about knowledge and knowing and their relation to learning. *Review of educational research*, *67*(1), 88-140.

Hofer, B. K. (2001). Personal epistemology research: Implications for learning and teaching. *Educational psychology review*, *13*(4), 353-383.

Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for *fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. Structural equation modeling: a multidisciplinary journal, 6*(1), 1-55.

Hughes, G. D. (2009). The impact of incorrect responses to reverse-coded survey items. *Research in the Schools*, *16*(2).

Johanson, G.A. and Osborn, C.J. (2000, April) Acquiescence as differential person functioning. Paper presented at the annual meeting of the American Education Research Association, New Orlean, LA (ERIC Document Reproduction Service No. ED 441 022).

Kahne, J., & Bowyer, B. (2017). Educating for democracy in a partisan age: Confronting the challenges of motivated reasoning and misinformation. *American Educational Research Journal*, *54*(1), 3-34.

Kammerer, Y., Amann, D. G., & Gerjets, P. (2015). When adults without university education search the Internet for health information: The roles of Internet-specific epistemic beliefs and a source evaluation intervention. *Computers in Human Behavior*, *48*, 297-309.

Kardash, C. M., & Scholes, R. J. (1996). Effects of preexisiting beliefs, epistemological beliefs, and need for cognition on interpretation of controversial issues. *Journal of Educational psychology*, *88*(2), 260.

Kavanagh, J., & Rich, M. D. (2018). *Truth decay: A threat to policymaking and democracy*. RAND.

Kendeou, P., Robinson, D. H., & McCrudden, M. T. (Eds.). (2019). *Misinformation and fake news in education*. IAP.

139

Kim, Y. S., Al Otaiba, S., Folsom, J. S., Greulich, L., & Puranik, C. (2014). Evaluating the
dimensionality of first-grade written composition. *Journal of Speech, Language, and
Hearing Research.*

Kim, Y. S., Al Otaiba, S., Wanzek, J., & Gatlin, B. (2015). Toward an understanding of
dimensions, predictors, and the gender gap in written composition. *Journal of
educational psychology, 107*(1), 79.

Kim, Y.-S. G., & Graham, S. (2021). Expanding the direct and indirect effects model of writing
(DIEW): Dynamic relations of component skills to various writing outcomes. Journal of
Educational Psychology. https://doi.org/10.1037/edu0000564

Kline, R. B. (2015). *Principles and practice of structural equation modeling*. Guilford
publications.

Kyle, K., & Crossley, S. (2016). The relationship between lexical sophistication and independent
and source-based writing. *Journal of Second Language Writing, 34*, 12-24.

Kuhn, D. (1999). A developmental model of critical thinking. *Educational researcher*, *28*(2), 16-
46.

Kuhn, D., Cheney, R., & Weinstock, M. (2000). The development of epistemological
understanding. *Cognitive development*, *15*(3), 309-328.

Kuhn, D. (2001). How do people know?. *Psychological science*, *12*(1), 1-8.

Kuhn, D., & Weinstock, M. (2002). What is epistemological thinking and does it matter?. BK
Hofer & PR Pintrich (Eds.) Personal epistemology: The psychology of beliefs about
knowledge and knowing (pp. 121-144) Mahwah.

Kuhn, D. (2019). Critical thinking as discourse. *Human Development*, *62*(3), 146-164.

Kuiper, E., Volman, M., & Terwel, J. (2005). The Web as an information resource in K–12

    education: Strategies for supporting students in searching and processing information.

    Review of educational research, *75*(3), 285-328.

Langer, J. A. (2011). *Envisioning knowledge: Building literacy in the academic disciplines.* New

    York, NY: Teachers College Press.

Lave, J. & Wenger, E. (1991). *Situated learning: Legitimate Peripheral Participation.*

    Cambridge: Cambridge University Press (Chpts. 1 & 2, pp. 29-58).

Lee, J., & Swan, K. (2013). Is the Common Core good for social studies? Yes, but…. *Social*

    Education, *77*(6), 327-330.

Lee, P., & Shemilt, D. (2003). A scaffold, not a cage: Progression and progression models in

    history. *Teaching history*, (113), 13.

Levine, P. (2014). The C3 Framework: One year later - an interview with Kathy Swan. *Social*

    Education, 78(4), 172-174, 178.

Lietz, P. (2008). Questionnaire design in attitude and opinion research: Current state of an art.

Limón, M., & Mason, L. (Eds.). (2002). *Reconsidering conceptual change: Issues in theory and*

    practice (pp. 115-135). Dordrecht: Kluwer Academic Publishers.

List, A., & Alexander, P. A. (2017). Cognitive affective engagement model of multiple source

    use. *Educational Psychologist*, *52*(3), 182-199.

List, A., Du, H., Wang, Y., & Lee, H. Y. (2019). Toward a typology of integration: Examining

    the documents model framework. *Contemporary Educational Psychology*, *58*, 228-242.

Lund, E. S., Bråten, I., Brandmo, C., Brante, E. W., & Strømsø, H. I. (2019). Direct and indirect

    effects of textual and individual factors on source-content integration when reading about

    a socio-scientific issue. *Reading and Writing*, *32*(2), 335-356.

MacArthur, C. A., Jennings, A., & Philippakos, Z. A. (2019). Which linguistic features predict

    quality of argumentative writing for college basic writers, and how do those features

    change with instruction?. *Reading and Writing*, *32*(6), 1553-1574.

MacCallum, R. C., Roznowski, M., & Necowitz, L. B. (1992). Model modifications in

    covariance structure analysis: the problem of capitalization on chance. *Psychological*

    bulletin, 111(3), 490.

Maggioni, L., VanSledright, B., & Alexander, P. A. (2009). Walking on the borders: A measure

    of epistemic cognition in history. *The Journal of Experimental Education*, *77*(3), 187-

    214.

Martin, D., Maldonado, S. I., Schneider, J., & Smith, M. (2011). A report on the state of history

    education: State policies and national programs. National History Education

    Clearinghouse.

Marsh, H. W. (1996). Positive and negative global self-esteem: A substantively meaningful

    distinction or artifactors?. *Journal of personality and social psychology, 70*(4), 810.

Mason, L., & Boscolo, P. (2004). Role of epistemological understanding and interest in

    interpreting a controversy and in topic-specific belief change. *Contemporary Educational*

    Psychology, *29*(2), 103-128.

Mason, L., & Bromme, R. (2010). Situating and relating epistemological beliefs into

    metacognition: studies on beliefs about knowledge and knowing. *Metacognition and*

    Learning, *5*(1), 1-6.

Mason, L., & Scirica, F. (2006). Prediction of students' argumentation skills about controversial

    topics by epistemological understanding. Learning and instruction, 16(5), 492-509.

Mason, L., Scirica, F., & Salvi, L. (2006). Effects of beliefs about meaning construction and task instructions on interpretation of narrative text. *Contemporary Educational Psychology*, *31*(4), 411-437.

McGrew, S., Breakstone, J., Ortega, T., Smith, M., & Wineburg, S. (2018). Can students evaluate online sources? Learning from assessments of civic online reasoning. *Theory & Research in Social Education*, *46*(2), 165-193.

McCrudden, M. T., & Sparks, P. C. (2014). Exploring the effect of task instructions on topic beliefs and topic belief justifications: A mixed methods study. *Contemporary Educational Psychology*, *39*(1), 1-11.

McCutchen, D. (2006). Cognitive factors in the development of children's writing. *Handbook of writing research*, *8*, 115-30.

Mo, Y., & Troia, G. A. (2017). Similarities and differences in constructs represented by US States' middle school writing tests and the 2007 National Assessment of Educational Progress writing assessment. *Assessing Writing, 33,* 48-67.

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

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

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

Monte-Sano, C. (2011). Beyond reading comprehension and summary: Learning to read and write in history by focusing on evidence, perspective, and interpretation. *Curriculum Inquiry*, *41*(2), 212-249.

Monte-Sano, C. (2012). What makes a good history essay? Assessing historical aspects of
      argumentative writing. *Social Education*, *76*(6), 294-298.

Moshman, D. (2018). Metacognitive theories revisited. *Educational Psychology Review*, *30*(2),
      599-606.

Muis, K. R., Bendixen, L. D., & Haerle, F. C. (2006). Domain-generality and domain-specificity
      in personal epistemology research: Philosophical and empirical reflections in the
      development of a theoretical framework. *Educational Psychology Review*, *18*(1), 3-54.

Muthén, L. K., & Muthén, B. O. (2017). Mplus user's guide (8th ed.). Los Angeles, CA: Muthén
      & Muthén. (Original work published 1998)

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

National Council for Social Studies. (2010). *National curriculum standards for social studies: A*
      framework for teaching, learning and assessment. Washington, DC: National Council for
      Social Studies

National Council for Social Studies. (2013). *Social studies for the next generation: Purposes,*
      practices, and implications of the college, career, and civic life (C3) framework for social
      studies state standards. Washington, DC: National Council for Social Studies.

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

National Writing Project. (2005, 2010). *The Analytic Writing Continuum: A comprehensive writing assessment system.* University of California, Berkeley; Berkeley, CA: National Writing Project.

Nokes, J. D., Dole, J. A., & Hacker, D. J. (2007). Teaching high school students to use heuristics while reading historical texts. *Journal of Educational Psychology*, *99*(3), 492.

Northwest Regional Educational Laboratory 6+1 Trait® Writing. 2011 Retrieved from http://educationnorthwest.org/traits.

Nunnally J. & Bernstein I. (1994). *Psychometric theory.* McGraw Hill, New York.

O'Hallaron, C. L. (2014). Supporting fifth-grade ELLs' argumentative writing development. *Written Communication, 31*(3), 304-331.

Olinghouse, N. G. (2008). Student-and instruction-level predictors of narrative writing in third-grade students. *Reading and Writing*, *21*(1), 3-26.

Olinghouse, N. G., Graham, S., & Gillespie, A. (2015). The relationship of discourse and topic knowledge to fifth graders' writing performance. *Journal of Educational Psychology*, *107*(2), 391.

Olson, C. B., Kim, J. S., Scarcella, R., Kramer, J., Pearson, M., van Dyk, D. A., Collins, P., & Land, R. E. (2012). Enhancing the interpretive reading and analytical writing of mainstreamed English learners in secondary school: Results from a randomized field trial using a cognitive strategies approach. *American Educational Research Journal, 49*(2), 323-355.

Olson, C. B., Matuchniak, T., Chung, H. Q., Stumpf, R., & Farkas, G. (2017). Reducing achievement gaps in academic writing for Latinos and English learners in Grades 7–12. *Journal of Educational Psychology, 109*(1), 1.

Olson, C. B., Woodworth, K., Arshan, N., Black, R., Chung, H. Q., D'Aoust, C., & Matuchniak, T. (2020). The pathway to academic success: Scaling up a text-based analytical writing intervention for Latinos and English learners in secondary school. *Journal of Educational Psychology, 112*(4), 701.

Paxton, R. J. (2002). The influence of author visibility on high school students solving a historical problem. *Cognition and Instruction*, *20*(2), 197-248.

Perfetti, C. A., Rouet, J. F., & Britt, M. A. (1999). Toward a theory of documents representation. *The construction of mental representations during reading*, *88108*.

Pessoa, S., Mitchell, T. D., & Reilly, B. (2019). Scaffolding the Writing of Argumentative Essays in History. *The History Teacher*, *52*(3), 411-440.

Reisman, A. (2012). The 'document-based lesson': Bringing disciplinary inquiry into high school history classrooms with adolescent struggling readers. *Journal of Curriculum Studies*, *44*(2), 233-264.

Reisman, A., Brimsek, E., & Hollywood, C. (2019). Assessment of historical analysis and argumentation (AHAA): A new measure of document-based historical thinking. *Cognition and Instruction*, *37*(4), 534-561.

Richter, T., & Maier, J. (2017). Comprehension of multiple documents with conflicting information: A two-step model of validation. *Educational psychologist*, *52*(3), 148-166.

Rouet, J. F., Britt, M. A., Mason, R. A., & Perfetti, C. A. (1996). Using multiple sources of evidence to reason about history. *Journal of Educational Psychology*, *88*(3), 478.

Rouet, J. F., & Britt, M. A. (2011). Relevance processes in multiple document comprehension. *Text relevance and learning from text*, 19-52.

Rouet, J.-F., Britt, M. A., & Durik, A. M. (2017). Resolv: Readers' representation of reading

    contexts and tasks. Educational Psychologist 52, 200-215.

    https://doi.org/10.1080/00461520.2017.1329015

Rogers, A., Stoeckel, J., & Sikali, E. (2013). National assessment of educational progress

    (NAEP) 2011 writing assessment restricted-use data files data companion, National

    Center for Education Statistics, Washington, D.C. (2013)

Roszkowski, M. J., & Soven, M. (2010). Shifting gears: Consequences of including two

    negatively worded items in the middle of a positively worded questionnaire. *Assessment*

    & Evaluation in Higher Education, *35*(1), 113-130.

Rukavina, I., & Daneman, M. (1996). Integration and its effect on acquiring knowledge about

    competing scientific theories for text. *Journal of Educational Psychology*, *88*(2), 272.

Scardamalia, M., & Bereiter, C. (1987). 4 Knowledge telling and knowledge transforming.

    Advances in Applied Psycholinguistics: *Volume 2, Reading, Writing, and Language*

    Learning, 2, 142.

Schommer, M. (1990). Effects of beliefs about the nature of knowledge on comprehension.

    Journal of educational psychology, *82*(3), 498.

Schommer, M. (1994). Synthesizing epistemological belief research: Tentative understandings

    and provocative confusions. *Educational psychology review*, *6*(4), 293-319.

Shanahan, T., & Shanahan, C. (2008). Teaching disciplinary literacy to adolescents: Rethinking

    content-area literacy. *Harvard educational review*, *78*(1), 40-59.

Shanahan, T., & Shanahan, C. (2012). What is disciplinary literacy and why does it matter?.

    T*opics in language disorders, 32*(1), 7-18.

Schipolowski, S., & Böhme, K. (2016). Assessment of writing ability in secondary education: comparison of analytic and holistic scoring systems for use in large-scale assessments. *L1 Educational Studies in Language and Literature*, *16*(Running Issue).

Schleppegrell, M.J. (2004). *The language of schooling: A functional linguistics perspective.* New York, NY: Routledge.

Schraw, G., & Moshman, D. (1995). Metacognitive theories. *Educational psychology review*, *7*(4), 351-371.

Schmitt, N., & Stults, D. M. (1986). Methodology review: Analysis of multitrait-multimethod matrices. *Applied Psychological Measurement*, *10*(1), 1-22.

Sinatra, G. M., Southerland, S. A., McConaughy, F., & Demastes, J. W. (2003). Intentions and beliefs in students' understanding and acceptance of biological evolution. *Journal of Research in Science Teaching: The Official Journal of the National Association for Research in Science Teaching*, *40*(5), 510-528.

Song, Y., & Ferretti, R. P. (2013). Teaching critical questions about argumentation through the revising process: Effects of strategy instruction on college students' argumentative essays. *Reading and Writing, 26*(1), 67-90.

Stadtler, M., & Bromme, R. (2014). The content–source integration model: A taxonomic description of how readers comprehend conflicting scientific information. *Processing inaccurate information: Theoretical and applied perspectives from cognitive science and the educational sciences*, 379-402.

Stanford History Education Group. (n.d.) *Montgomery Bus Boycott*. https://sheg.stanford.edu/history-lessons/montgomery-bus-boycott

Steiss, J., Krishnan, J., Kim, Y. S. G., & Olson, C. B. (2022). Dimensions of text-based analytical writing of secondary students. *Assessing Writing*, *51*, 100600.

Stoel, G. L., van Drie, J. P., & Van Boxtel, C. A. (2017). The effects of explicit teaching of strategies, second-order concepts, and epistemological underpinnings on students' ability to reason causally in history. *Journal of Educational Psychology*, *109*(3), 321.

Strømsø, H. I., Bråten, I., & Samuelstuen, M. S. (2008). Dimensions of topic-specific epistemological beliefs as predictors of multiple text understanding. *Learning and Instruction*, *18*(6), 513-527.

Strømsø, H. I., Bråten, I., & Britt, M. A. (2011). Do students' beliefs about knowledge and knowing predict their judgement of texts' trustworthiness?. *Educational Psychology*, *31*(2), 177-206.

Strømsø, H. I., Bråten, I., Britt, M. A., & Ferguson, L. E. (2013). Spontaneous sourcing among students reading multiple documents. *Cognition and Instruction*, *31*(2), 176-203.

Taber, C. S., & Lodge, M. (2006). Motivated skepticism in the evaluation of political beliefs. *American journal of political science*, *50*(3), 755-769.

Toulmin, S. (1958). The uses of argument. Cambridge, England: Cambridge University Press.

Troia, G. A., Shen, M., & Brandon, D. L. (2019). Multidimensional levels of language writing measures in grades four to six. *Written Communication, 36*(2), 231-266.

Van Drie, J., & Van Boxtel, C. (2008). Historical reasoning: Towards a framework for analyzing students' reasoning about the past. *Educational Psychology Review*, *20*(2), 87-110.

VanSledright, B. (2002). Confronting history's interpretive paradox while teaching fifth graders to investigate the past. *American Educational Research Journal*, *39*(4), 1089-1115.

VanSledright, B., & Maggioni, L. (2016). Epistemic cognition in history. In *Handbook of epistemic cognition* (pp. 140-158). Routledge.

Voss, J. F., Wiley, J., & Carretero, M. (1995). Acquiring intellectual skills. *Annual Review of Psychology*, *46*(1), 155-181.

Voss, J., Wiley, J., & Kennet, J. (1998). Student perceptions of history and historical concepts. *International review of history education*, *2*, 307-330.

Wagner, R. K., Puranik, C. S., Foorman, B., Foster, E., Tschinkel, E., & Kantor, P. T. (2011). Modeling the development of written language. *Reading and Writing: An Interdisciplinary Journal, 24*, 203–220.

Wang, E., Matsumura, L. C., & Correnti, R. (2018). Student writing accepted as high-quality responses to analytic text-based writing tasks. *The Elementary School Journal,* 118(3), 357-383.

Weems, G. H., Onwuegbuzie, A. J., & Lustig, D. (2003). Profiles of respondents who respond inconsistently to positively-and negatively-worded items on rating scales. *Evaluation & Research in Education, 17*(1), 45-60.

Wiley, J., Griffin, T. D., Steffens, B., & Britt, M. A. (2020). Epistemic beliefs about the value of integrating information across multiple documents in history. *Learning and Instruction*, *65*, 101266.

Wiley, J., Griffin, T. D., Steffens, B., & Britt, M. A. (2020). Epistemic beliefs about the value of integrating information across multiple documents in history. *Learning and Instruction*, *65*, 101266.

Wilson, J., Roscoe, R., & Ahmed, Y. (2017). Automated formative writing assessment using a levels of language framework. *Assessing Writing, 34*, 16-36.

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

Wineburg, S. (Ed.). (2001). Historical thinking and other unnatural acts: Charting the future of teaching the past. Philadelphia, PA: Temple University Press.

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

Wolfe, M. B., & Goldman, S. R. (2005). Relations between adolescents' text processing and reasoning. *Cognition and Instruction*, *23*(4), 467-502.

Wolfe, C. R. (2012). Individual differences in the "myside bias" in reasoning and written argumentation. *Written Communication*, *29*(4), 477-501.

Yli-Panula, E., Laakkonen, E., & Vauras, M. (2021). High-school students' topic-specific epistemic beliefs about climate change: An assessment-related study. *Education Sciences*, *11*(8), 440.

Zavala, J., & Kuhn, D. (2017). Solitary discourse is a productive activity. *Psychological science*, *28*(5), 578-586.

Appendix A. Source-based argument writing prompts

Why did the Delano Grape Strike and Boycott Succeed?

Day 1

Essential Question:

Why did the Delano Grape Strike and Boycott succeed?

- This is a **two-day** activity that is intended to see what type of argument writing students do using these sources.

- On **Day 1**, you will have 50-minutes.
  - Listen to me read the sources OR read the sources yourself
  - Pause after each source to highlight important information and complete your thinking questions.

- On **Day 2**, you will have 50-minutes.
  - Write an argument responding to the essential question
  - Use evidence from the sources

Source 1. Strike and Boycott Timeline
Source 2. Roberto Bustos Interview
Source 3. César Chavez Letter
Source 4. Dolores Huerta Speech

Note: To help you write an **argument** about why the strike and boycott succeeded, first read these sources like a historian. Ask yourself:

- Who wrote the source? When? Where? Why?
- How does the time and place that the source was created impact its content?
- Where do sources agree? Where do the sources disagree with each other? Why?

You may watch the video to listen to someone read the information from the PDF.

Why did the Delano Grape Strike and Boycott Succeed?

Open the PDF and follow along and annotate or if you would like to read them independently, open the PDF and annotate on your own.

Headnote: To fight injustice, individuals and groups have used different strategies to speak out and address social issues they see as important. One strategy to produce change is a **strike**—a form of protest where workers refuse to work due to unfair pay. Another strategy is a **boycott**—a form of protest where people do not buy goods or services due to moral, social, or political reasons.

Background: In the 1960s, farm workers in California began to organize and protest poor working conditions and low pay. A group of workers in Delano, CA began a **strike** in 1965. Later, these workers joined another group to create the **United Farm Workers** Union (UFW). They aimed to increase pay and improve working conditions.

The UFW organized a workers' strike, a grape boycott, and community marches. These actions helped the movement gain national attention.

In July 1970, an agreement was reached between the workers and the many major grape growers in California. The agreement, which affected more than 10,000 farm workers, included an increase in pay and improved working conditions for workers in Delano and other areas across California.

The images are from a TIME Magazine article in 1969. Top image: People encourage shoppers to boycott grapes. Bottom image: Farm workers and strikers march 340 miles from Delano to Sacramento, California to bring attention to the struggle of grape workers.

153

Why did the Delano Grape Strike and Boycott Succeed?

Source 1 Strike and Boycott Timeline

Headnote: This timeline shows events from when the Delano grape strike began in 1965 to when an agreement was signed between Delano grape growers and their workers in 1970.

Date	Event	YOUR THINKING
Sept 1965	Filipino American grape workers strike in Delano, California. The mostly Latino workers, led by César Chavez, join the strike.	According to the source, what helped the strike and boycott succeed?
March 17, 1966	Roberto Bustos and César Chavez lead farm workers on a 340 mile march from Delano to Sacramento to bring attention to the struggle of grape workers. Grape workers in Delano reach an agreement with their employer, but the strike continues against other growers.	
Aug 1966	Latino and Filipino Unions join together to form the United Farm Workers Union (UFW) led by César Chavez. The UFW cannot reach an agreement with grape growers.	What evidence from the source makes you think this?
1966-1967	Dolores Huerta leads striking grape workers and supporters in a national boycott of California grapes. Hundreds of grape strikers traveled across the U.S., telling their stories to gain support.	
Feb 1968	César Chavez speaks on TV and writes letters asking people to stop buying all California grapes.	Explain why this is important for the success of the strike and boycott.
May 10, 1969	Dolores Huerta speaks for grape workers on international boycott day, asking the U.S. supporters for their continued support.	
Sept 1970	UFW signs long-term agreements with the Delano grape growers and other growers in California. This agreement includes better pay and treatment for grape workers. The grape workers' strike and boycott ends.	How useful is this source for thinking about why the strike and boycott succeeded? Is the evidence reliable, trustworthy, and relevant?

Source Note: The timeline comes from the website: The Fight in the Fields: César Chavez and the Farm Workers' Struggle (http://www.fightinthefields.net/timeline.html). The site describes the Farm workers' movement and sells books and videos about the movement.

154

Why did the Delano Grape Strike and Boycott Succeed?

Source 2 Roberto Bustos Interview

Headnote: Roberto Bustos spent his early life as a farm worker. He later became a part of the United Farm Workers Union (UFW). In 1966, Bustos organized the 340 mile march from Delano to Sacramento to bring attention to the struggle of grape workers. He answers questions by a reporter named Juan Esparza Loera.

	YOUR THINKING
Loera: What are your memories of the 1966 march to Sacramento?	
Bustos: César Chavez came to us and said 'Guys, what do you think if we march to Sacramento? People are not picking grapes, they're striking. Let's see what the governor of California can do for us. The growers are not answering our letters.'	According to the source, what helped the strike and boycott succeed?
We said we'll go. We figured we would drive 3 hours to Sacramento, but then César said he was walking. We thought this guy was crazy! No farm workers had walked to the state Capitol.	
César said I would be in charge of the march. He said, 'Roberto, people listen to you.' He said 'We will march in other farm worker communities to tell people why we're doing this and to join us. And, tell them not to break the workers' strike.'	What evidence from the source makes you think this?
We stopped at 53 towns and explained why we were marching.	
Loera: What was your first impression of Dolores Huerta and César Chávez'?	Explain why this is important for the success of the strike and boycott.
Bustos: It was not until Dolores went out in the fields that I found out about her and who she was. She is a great lady. I call her now "The Jefa" (The Boss).	
When I first met César, I thought, he knows what he's talking about: Work, fair pay and treatment. He's talking about my family. He knows our experiences.	How useful is this source for thinking about why the strike and boycott succeeded? Is the evidence reliable, trustworthy, and relevant?

Source Note: This interview was in the newspaper *Vida en el Valle*, published October 5th, 2017. The person being interviewed, Roberto Bustos, organized the first Delano-to-Sacramento march in 1966 for the United Farm Workers.

Why did the Delano Grape Strike and Boycott Succeed?
Source 3 César Chavez Letter

Headnote: This letter was written by César Chavez. It asks people to tell their grocery stores to stop selling California grapes until workers receive better wages. The UFW helped send this letter out to thousands of people.

	YOUR THINKING
Dear Los Angeles Friend,	
Another Christmas season finds us grape workers striking and boycotting grapes all over the United States, instead of celebrating Christmas as you are.	According to the source, what helped the strike and boycott succeed?
This is a season when men pay special attention to the needs of their brothers and we are grateful for the support that has come from some of our Los Angeles friends.	
	What evidence from the source makes you think this?
We only want one thing: to escape from poverty by getting the grape growers to provide good pay and working conditions. Meanwhile the only way for us to gain these rights is by the strike and the boycott.	
But we cannot win without your help. During the holiday season, please remember the farm worker, who harvests your food. DON'T BUY GRAPES! Tell your friends and coworkers to NOT buy grapes.	
If you want to make the Christmas merrier for grape workers, you can do even more. You can join our boycott. You and your friends can visit every store in your neighborhood, and tell them you won't shop there until they remove the grapes!	Explain why this is important for the success of the strike and boycott.
Merry Christmas,	
César Chavez	
	How useful is this source for thinking about why the strike and boycott succeeded? Is the evidence reliable, trustworthy, and relevant?
p.s. Please don't buy grapes!	

Source Note: This open letter was written by Chavez in 1968. It was published and thousands of copies were likely made and spread throughout the Los Angeles community to create support for the UFW and the grape boycott. The original document was *Retrieved from University of California Berkeley Digital Library.*

156

Headnote: Dolores Huerta was a leader of grape workers. Huerta frequently negotiated with lawmakers to win rights for grape workers. This is a speech that Huerta gave in 1969 that explained the grape workers' concerns and why they wanted shoppers to continue to boycott grapes.

	YOUR THINKING
We, the striking grape workers of California, join with U.S. shoppers to work towards freedom. As we boycott, we think about the steps that got us here. After marching from Delano to Sacramento, we have talked with people in other cities, helping them understand our problems with the grape growers.	According to the source, what helped the strike and boycott succeed?
We have been farm workers for hundreds of years. Mexicans, Filipinos, Africans and others. If our actions lead to the rights we demand, to good pay, to humane working conditions, to protection from pesticides, and the right to negotiate pay, thousands of American farm workers will follow us.	What evidence from the source makes you think this?
We have been farm workers for hundreds of years and strikers for four years. It was four years ago that we threw down our work tools in protest of poor working conditions. We went and stood tall outside the vineyards where we had bent over for years. But national labor laws left us unprotected and grape growers refused to offer good pay and improve working conditions. Yet we knew the day must come when they would talk to us as equals.	
Because the strike and marches were not enough, we had to boycott. We did not choose the grape boycott, but we had chosen to leave our poverty behind.	Explain why this is important for the success of the strike and boycott.
We marched alone at the beginning, but today we count people from all over the country in our number. The shoppers who support our cause are boycotting by not buying grapes! And they won't shop at stores that sell grapes during the boycott, just as we withhold our labor from the growers until our lives improve.	
The grapes grow sweet and heavy on the vines, but they will have to wait while we reach out first for our freedom. The time is ripe for our liberation.	How useful is this source for thinking about why the strike and boycott succeeded? Is the evidence reliable, trustworthy, and relevant?

Source Note: Dolores Huerta gave this speech, called "Proclamation of the Delano Grape Workers" on International Boycott Day on May 5, 1969. She participated in strikes, boycotts and worked with lawmakers to make changes to help workers.

Your Task: Your school is having a discussion about how people organize and act to fight injustice and produce social change. Your school has asked student-experts to write about different boycotts and the most significant reasons for their successes. You have been given a number of important sources that you read like a historian to better understand the Delano Grape Strike and Boycott. Understanding why the boycott succeeded will enable you to apply these strategies to fight injustice today.

Writing Prompt: Write **an argument** that responds to the following question:
Why did the Delano Grape Strike and Boycott succeed?

In the **Introduction**
- ❖ inform your reader by explaining the **background** of the Delano grape strike and boycott, and
- ❖ write a **clear claim that responds to the essential question:** Why did the Delano Grape Strike and Boycott succeed?

In the **body**
- ❖ Select and analyze one significant **reason for the success** of the strike and boycott. Support your idea with **evidence from the sources**. Share your reasoning on why this led to the strike and boycott's success.
- ❖ Next, select and analyze another significant **reason for the success** of the strike and boycott. Support your idea with **evidence from the sources**. Share your reasoning on why this led to the strike and boycott's success.
- ❖ Evaluate the credibility of the sources. Explain which of your reasons was **most important** for the success of the strike and boycott. Acknowledge why someone else might argue that a different reason was most important. Defend your argument with *your* strongest evidence.

In the **conclusion**
- ❖ Remind your reader of your reasons for **why the Delano grape strike and boycott succeeded**
- ❖ Provide a statement on what we can learn from the success of the Delano strike and boycott and how we can apply these strategies to fight injustice today

Why did the Montgomery Bus Boycott Succeed?

Day 1

Essential Question:
Why did the Montgomery Bus Boycott succeed?

- This is a **two-day** activity that is intended to see what type of argument writing students do using these sources

- On **Day 1**, you will have 50-minutes.
  - Listen to me read the sources OR read the sources yourself
  - Pause after each source to highlight important information and complete your thinking questions

- On **Day 2**, you will have 50-minutes.
  - Write an argument responding to the essential question
  - Use evidence from the sources

Source 1. Bus Boycott Timeline
Source 2. Jo Ann Robinson's Memoir
Source 3. Bayard Rustin's Diary
Source 4. Rosa Parks Biography

Note: To help you write an argument about why the boycott succeeded, first read these sources like a historian. Ask yourself:

- Who wrote the source? When? Where? Why?
- How does the time and place that the source was created impact its content?
- Where do sources agree? Where do the sources disagree with each other? Why?

Why did the Montgomery Bus Boycott Succeed?

Headnote: To fight injustice, individuals and groups have used public platforms to express their views and act to address social issues they see as important. One strategy to produce change is a **boycott**—a protest where people do not buy goods or services due to moral, social, or political reasons.

Background: Like many U.S. cities in 1955, Montgomery, Alabama's city buses were segregated. This meant that the first ten rows of seats were reserved for White people, while Black people were forced to sit or stand in the rear of the bus. This was one of the many discriminatory laws that negatively affected the lives of Black people during the 1950s.

The Women's Political Council (WPC), Black leaders, and other groups formed the Montgomery Improvement Association (MIA) on December 5, 1955 to end bus segregation and to move towards equal rights for Black citizens. The MIA's weekly public meetings were attended by thousands of people. They helped organize the Montgomery Bus Boycott, which occurred after the arrest of Rosa Parks. For months, the city buses were almost empty because large numbers of Black customers participated in the Montgomery Bus Boycott. This ended in the U.S. Supreme Court ruling that segregation on public buses is unconstitutional and illegal.

Above: Images from a website BlackPast.org. People are walking in Montgomery, Alabama instead of riding the bus.

Why did the Montgomery Bus Boycott Succeed?

Source 1 Bus Boycott Timeline

Headnote: This timeline shows events from before the boycott to the Supreme Court ruling that segregation on public buses is unconstitutional and illegal.

Date	Bus Boycott Event Timeline	YOUR THINKING
March 1955	Black leaders in Montgomery, including Rosa Parks, Martin Luther King Jr., and The Women's Political Council (WPC) meet with city officials to recommend changes for the Montgomery bus system.	According to the source, what helped the boycott succeed?
Dec 1, 1955	Rosa Parks is arrested for refusing to give up her seat to a White passenger.	What evidence from the source makes you think this?
Dec 5, 1955	The WPC distributed thousands of handouts to Black people urging them to stay off the buses in a one-day boycott.	
Dec 5, 1955	After a successful one-day boycott, the Montgomery Improvement Association (MIA) is founded to continue the boycott and stake more actions for change in Montgomery	Explain why this is important for the success of the boycott.
Dec 13, 1955	The MIA, led by Dr. Martin Luther King, Jr. and Rufus Lewis, creates a carpool system to support citizens taking part in the boycott.	
June 5, 1956	The Supreme Court's orders of injunction against segregation on city buses are delivered to the Montgomery City Hall.	How useful is this source for thinking about why the boycott succeeded? Is the evidence reliable, trustworthy, and relevant?
Dec 21, 1956	Montgomery's buses are officially desegregated. The Montgomery Improvement Association (MIA) ends the boycott.	

Source Note: A Timeline of the Montgomery Bus Boycott was published on https://www.beaconbroadside.com, a website for a publishing company.

Why did the Montgomery Bus Boycott Succeed?

Source 2 Jo Ann Robinson's Memoir

Headnote: After unsuccessful meetings with Montgomery officials, Black leaders sought other ways to create change. After Rosa Parks was arrested for refusing to give up her bus seat to a White passenger, Black leaders decided to boycott city buses. Jo Ann Robinson, the president of the Women's Political Council, explains how she and others produced and distributed a handout to thousands of Black people urging them to stay off the buses on Monday morning, December 5, 1955.

	YOUR THINKING
I sat down, wrote a message, and then called a good friend and colleague... who had access to the college's printing equipment. When I told him that the Women's Political Council (WPC) was staging a boycott and needed to print the handouts, he told me that he too had suffered embarrassment on the city buses.... Along with two of my most trusted senior students, we quickly agreed to meet, in the middle of the night, at the college's copy room. We were able to get three handouts to a page in order to make thousands of handouts we knew would be needed. By 4 a.m. Friday, the handouts had been made....	According to the source, what helped the boycott succeed?
Between 4 and 7 a.m., the two students and I mapped out distribution routes for the handouts. Some of the WPC officers previously had discussed how and where to deliver thousands of paper handouts announcing a boycott, and those plans were helpful to me....	What evidence from the source makes you think this?
After class, my two students and I quickly finalized our plans for distributing the thousands of handouts so that one would reach every Black home in Montgomery. I took out the WPC membership roster and called everyone.... I told them about the need to distribute the handouts and then helped distribute them to the community....	Explain why this is important for the success of the boycott.
We spent the day dropping off thousands of handouts. Some were dropped off at schools, some were dropped off at business places, storefronts, salons, factories, barber shops, and every other available place. Workers would pass along handouts both to other employees as well as to customers.... By 2 o'clock, thousands of handouts had been given to many people. Practically every Black man, woman, and child in Montgomery knew the plan and was passing the word along...	How useful is this source for thinking about why the boycott succeeded? Is the evidence reliable, trustworthy, and relevant?

Source Note: This excerpt comes from Jo Ann Robinson's memoir: *The Montgomery Bus Boycott and the Women Who Started It.* Jo Ann Robinson was present during the boycott and a member of the WPC.

Why did the Montgomery Bus Boycott Succeed?

Source 3 Bayard Rustin's Diary

Headnote: Without the use of city buses, many Black citizens had to find other means of transportation, such as car pools. Bayard Rustin, a Black, gay, civil rights activist, traveled to Montgomery to advise Dr. King and the Montgomery Improvement Association and support the bus boycott. He kept a diary of what he found.

February 24	YOUR THINKING
42,000 Black citizens have not ridden the busses since December 5. On December 6, the police began to harass, intimidate, and arrest Black taxi drivers who were helping get these people to work. It thus became necessary for the Black leaders to find an alternative—the car pool. They set up 23 locations where people could gather to wait for free transportation.	According to the source, what helped the boycott succeed?
This morning Rufus Lewis, the director of the car pool and member of the MIA, invited me to attend the meeting of the drivers. On the way, he explained that there are three methods in addition to the car pool, for moving the Black population:	What evidence from the source makes you think this?
1) Hitch-hiking.	Explain why this is important for the success of the boycott.
2) The transportation of servants by White housewives.	
3) Walking.	
Later he introduced me to two men, one of whom has walked 7 miles and the other 14 miles, every day since December 5.	How useful is this source for thinking about why the boycott succeeded? Is the evidence reliable, trustworthy, and relevant?
"The success of the car pool is at the heart of the movement," Lewis said at the meeting. "It must not be stopped."	
I wondered what the response of the drivers would be, since 28 of them had just been arrested on charges of conspiring to destroy the bus company. One by one, they pledged that, if necessary, they would be arrested again and again.	

Source Note: This excerpt is from Bayard Rustin's *Montgomery Diary*, February 24, 1956. He travelled to Montgomery and described the ongoing boycott, which began months before he arrived.

Why did the Montgomery Bus Boycott Succeed?

Source 4 Rosa Parks Biography

Headnote: By 1955, Rosa Parks was an established leader in the Civil Rights Movement. Parks not only showed resistance by refusing to move from her bus seat, she also helped plan the Montgomery Bus Boycott. Many have tried to diminish Parks' role in the boycott by depicting her as a seamstress who simply did not want to move because she was tired. Parks denied the claim and years later revealed her true motivation which is included in the following article.

Rosa Parks (1913-2005)	YOUR THINKING
On December 1, 1955, Parks boarded a bus in Montgomery, Alabama. Instead of going to the back of the bus, where Black passengers sat, she sat in the front. When the bus started to fill up with White passengers, the driver asked Parks to move. She refused. Her resistance set in motion one of the largest social movements in history, the Montgomery Bus Boycott.	According to the source, what helped the boycott succeed?
Rosa Parks was born in 1913 in Tuskegee, Alabama. As a young woman, Parks enrolled at Alabama State Teachers College, a college for Black students. Growing up in the segregated South, Parks was frequently confronted with racial discrimination and violence. She became active in the Civil Rights Movement at a young age.	What evidence from the source makes you think this?
Parks, with her husband, worked with many social justice organizations. Eventually, Parks was elected secretary of the Montgomery chapter of the National Association for the Advancement of Colored People (NAACP). She said,	Explain why this is important for the success of the boycott.
"People always say that I didn't give up my seat because I was tired, but that isn't true. I was not tired physically, or no more tired than I usually was at the end of a working day. I was not old, although some people have an image of me as being old. I was forty-two. No, the only tired I was, was tired of giving in." Parks' courageous act and the subsequent Montgomery Bus Boycott led to the integration of public transportation in Montgomery. Her actions were not without consequence. She was jailed for refusing to give up her seat and lost her job for participating in the boycott.	How useful is this source for thinking about why the boycott succeeded? Is the evidence reliable, trustworthy, and relevant?

Source Note: "Rosa Parks (1913-2005)" was written by Arlisha Norwood and published by the National Women's History Museum in 2017.

Your Task: Your school is having a discussion about how people organize and act to fight injustice and produce social change. Your school has asked student-experts to write about different boycotts and the most significant reasons for their successes. You have been given a number of important sources that you read like a historian to better understand the 1955 Montgomery Bus Boycott. Understanding why the boycott succeeded will enable you to apply these strategies to fight injustice today.

Writing Prompt: Write **an argument** that responds to the following question: <mark>**Why did the Montgomery Bus Boycott succeed?**</mark>

In the **Introduction**
❖ inform your reader by explaining the **background** of the Montgomery Bus Boycott in 1955, and
❖ write a **clear claim that responds to the essential question:** Why did the Montgomery Bus Boycott succeed?

In the **body**
❖ Select and analyze one significant **reason for the success** of the bus boycott. Support your idea with **evidence from the sources**. Share your reasoning on why this led to the boycott's success.
❖ Next, select and analyze another significant **reason for the success** of the bus boycott. Support your idea with **evidence from the sources**. Share your reasoning on why this led to the boycott's success.
❖ Evaluate the credibility of the sources. Explain which of your reasons was **most important** for the success of the bus boycott. Acknowledge why someone else might argue that a different reason was most important. Defend your argument with *your* strongest evidence.

In the **conclusion**
❖ Remind your reader of your reasons for **why the Montgomery Bus Boycott succeeded**
❖ Provide a statement on what we can learn from the success of the Montgomery Bus Boycott and how we can apply these strategies to fight injustice today

Appendix B. Holistic writing rubric

Holistic Rubric for Source-Based Argument Writing in History

6 Exceptional Achievement

- Presents a clear, compelling, and accurate argument that addresses all requirements of the prompt
- Supports claims with relevant and sufficient evidence and compelling reasoning that connects evidence to claims
- Integrates sufficient, appropriate evidence from multiple sources and attributes evidence to sources, citing the title, author, and/or genre
- Analyzes relevant source information, such as context and perspective, and explains how sources relate
- Effectively presents and refutes/addresses an alternative perspective with sufficient evidence and reasoning
- Writing is well organized with a strong introduction, body, and conclusion and transitions to create coherence
- Demonstrates effective and varied sentence fluency with little to no errors in writing conventions
- Uses sophisticated language and academic tone

5 Commendable Achievement

- Presents a clear and accurate argument that addresses most requirements of the prompt
- Supports claims with mostly relevant and sufficient evidence and reasoning that connects evidence to claims
- Integrates appropriate evidence from multiple sources and mostly attributes evidence to sources
- Discusses source information, such as context and perspective, and may or may not explain how sources relate
- Competently presents and refutes/addresses an alternative perspective with evidence and reasoning
- Writing is organized, with a clear introduction, body, conclusion, and some transitions to create coherence
- Demonstrates effective sentence fluency that has limited errors in writing conventions
- Uses mostly sophisticated language and academic tone

4 Generally Competent Achievement

- Presents a simple argument that addresses some requirements of the prompt
- Supports most claims with evidence and reasoning that makes connections to claims
- Integrates evidence from at least 2 sources; minimally attributes evidence to sources
- Discusses some source information, such as context and perspective, for most of the sources used, and may or may not explain how sources relate
- Presents and refutes/addresses an alternative perspective with limited evidence and reasoning
- Writing is sufficiently organized with an opening, body, conclusion, and some transitions; the central claim may not appear in the opening
- Demonstrates competent sentence fluency that may have some errors in writing conventions
- Uses some sophisticated language and mostly academic tone

3 Developing Evidence of Achievement

- Presents a partially developed or unclear argument that superficially addresses some requirements of the prompt
- Partially supports claims, but lacks sufficient evidence or clear and logical reasoning that makes connections to claims
- Integrates insufficient evidence and may not attribute evidence to sources
- Provides some source information but presents limited or inaccurate analysis of source information, such as context and perspective
- Superficially presents and refutes/addresses an alternative perspective
- Writing is somewhat organized, may be lacking an opening, body, conclusion; lacks transitions to maintain coherence
- Demonstrates developing sentence fluency that has many errors in writing conventions that does not interfere with the reader's understanding
- Uses simple language and inconsistent academic tone

2 Minimal Evidence of Achievement

- Presents unclear or inaccurate argument, but addresses at least one requirement of the prompt
- Claims are unclear or unsupported; evidence and reasoning are mostly insufficient or illogical
- Includes source material that lacks relevance, integration, and attribution
- Lacks source information such as context and perspective, but shows some understanding of the time, place, and events discussed in sources
- Presents limited or inaccurate alternative perspective
- Writing is unorganized, brief, or underdeveloped with minimal coherence
- Uses simple language, writing style, and tone with many errors in writing conventions that interfere with reader's understanding
- Demonstrates ineffective sentence fluency; contains many errors in writing conventions that interfere with the reader's understanding
- Uses simple language and informal tone

1 No Evidence of Achievement

- Writing fails to address the prompt or makes a flawed attempt that is brief or incoherent
- Claims, evidence, and reasoning are absent or unrelated to prompt
- Includes source material that is almost entirely copied or presents summary
- Lacks source information such as context and perspective and does not show understanding of the time, place, and events discussed in sources
- Lacks an alternative perspective
- Writing has no organization or coherence
- Demonstrates ineffective sentence fluency; contains serious errors in writing conventions that interfere with the reader's understanding
- Uses flawed or inappropriate language and informal tone