Effects of Instructional Conditions on Comprehension from Multiple Sources in History and Science

Jennifer Wiley[1], Allison J. Jaeger[2], and Thomas D. Griffin[1]

[1]University of Illinois at Chicago and [2]Temple University

Author Note

The authors were supported by grants from the National Science Foundation (DUE 1535299) and the Institute of Education Sciences (R305A160008) during the preparation of this chapter. All opinions expressed herein are those of the authors and do not necessarily reflect those of the funding agencies.

Final accepted manuscript version, July 11, 2017

This manuscript can be cited as:

Wiley, J., Jaeger, A. J., & Griffin, T. D. (2018). Effects of task instructions on comprehension from multiple sources in history and science. In J. L. G. Braasch, I. Bråten, & M. T. McCrudden (Eds.) *Handbook of multiple source use* (pp. 341-361). New York, NY: Routledge.

Effects Instructional Conditions on Comprehension from Multiple Sources in History and Science

The purpose of this chapter is to review theory and research on the effects of instructional manipulations and conditions on comprehension of events or phenomenon in science and history from multiple sources. Throughout this chapter the terms text, source, or document are generally used interchangeably to refer to a body of information that is presented in some way that denotes it as an entity distinct from other bodies of information that are presented alongside it. This could be done by having bodies of information on separate physical pages, within separate windows of a browser, accessed via different web links, etc. Different studies have employed stimuli ranging from websites, excerpts of magazine or journal articles, news briefs, or passages from books; which may or may not explicitly include authorship or other referential information. They can include text, tables, or graphics, exclusively or in combination.

Within the domain of history, there is a long tradition of using multiple sources as part of the inquiry process, with the goal for such inquiry being the development of an account of how or why things may have happened in the past (Wiley & Ash, 2005; Wineburg, 1991). This multiple-source inquiry process is the central activity for many professional historians, and primary evidence for various historical accounts of events typically comes from other existing documents. The multiple-source inquiry process is also part of disciplinary instruction in History, most prominently as part of Advanced Placement (AP) courses in which students learn to respond to document-based questions. In contrast, the central task of professional scientists is answering questions via their own direct observations of events within the context of an experiment. Integrating information from multiple documents is essential to how scientists determine what is already known, what hypotheses are in need of testing, and what methods to

use in their own research. However, these earlier stages of the scientific process involving multiple-document inquiry have received far less attention as part of disciplinary instruction in science. Typically, students are merely presented with the final conclusions from prior scientific work, with more focus upon experimental data collection processes. With the emergence of the Internet, the relevance of comprehending from multiple sources has become ever more pertinent across a wide variety of topics and disciplines. Today, the most common method that lay people use to gain an understanding of historical events or scientific phenomena is via Internet searches, displacing the practice of using a bound and printed encyclopedia set or textbook as a reference.

Starting with early research on multiple source comprehension that primarily emerged from work in history, researchers have explored several types of instructional manipulations including altering the features of the inquiry task that is given (such as being asked to write a narrative or an argument); changing features of the task environment (such as the format of the source documents or features of the document set); and varying the instructional context (such as having students engage in a particular activity or training prior to engaging in a multiple-source inquiry task). These three broad categories continue to represent the main types of manipulations that have been studied in the literature. Further, this literature now includes both studies that have explored the comprehension of historical events from multiple-source inquiry tasks, as well as studies that have explored the development of understanding about scientific phenomena. To outline the remainder of this chapter, the next section provides an overview discussing the kinds of processes that are theoretically involved in multiple-source comprehension, and articulates the challenges that readers face when they attempt to engage in multiple-source inquiry tasks. Then, the main section of the chapter provides a summary of empirical research that has attempted to explore multiple source comprehension processes using manipulations of the inquiry task, the

task environment, and the instructional context. Research from studies in both history and science are included in an attempt to outline possible similarities and differences. A final section concludes the chapter with a theoretical synthesis derived from the empirical review, a discussion of future directions for research, and practical implications of this work.

Theoretical Background

Engaging in inquiry using multiple sources is a complex activity because it requires all the processes necessary for comprehending individual informational sources, plus an additional set of processes that become particularly important when readers are confronted with information from more than one source. According to theories of text comprehension, understanding even a single informational text requires the construction of multiple levels of representation (Graesser, Singer, & Trabasso, 1994; Kintsch, 1998). At the most basic level, a reader creates a *surface* representation, which generally consists of a fleeting episodic trace capturing the exact words and format of the source. At the next level of processing, the reader attempts to develop a *textbase* representation. This is essentially a propositional representation of the ideas presented in each clause or sentence. Basic word and sentence-level reading processes contribute to the construction of the *textbase*. In addition, as part of the comprehension of informational text, the reader must attempt to develop yet another level of representation, referred to as the *situation model*. On this level, the reader attempts to connect ideas between the sentences that appear in the text, and connect ideas in the text with prior knowledge, to develop a coherent understanding of the content that is being described.

When readers are confronted with more than a single source from which to obtain information, then the situation becomes more complicated. Comprehending events or phenomena from multiple sources instead of a single source necessitates representation of information about

the nature or origin of the various documents (e.g., who wrote it, for what purpose was it written), and relations among the documents (e.g. the presence of corroborating or conflicting statements). In these situations, the term 'source' is commonly used in a more restricted manner than the way it was defined at the outset of this chapter. Specifically, the term 'source' can also refer to "information about individuals and organizations that create and publish textual content, including information about when, where, and for what purpose the content is created and published" (Bråten & Braasch, 2018/this volume). According to the MD-Trace framework (Rouet & Britt, 2011), this information is captured by the *intertext model*.

Further, there needs to be a level of representation that reflects the understanding derived from integrating information across multiple sources. Beyond developing representations of each individual source, the reader needs to develop a *documents model* (Perfetti, Rouet, & Britt, 1999) or an *integrated model* to serve as the representation of the readers' understanding of the situation or phenomenon being described (Britt & Rouet, 2012). In a multiple source context, it is this level that best represents a reader's understanding of the content, or the mental model that has been constructed about the phenomenon that is the focus of inquiry (Wiley & Voss, 1999; Wiley et al., 2009).

The development of this understanding is also critically influenced by the reader's interpretation of the task and the processing they should engage in to achieve it (Wiley, Griffin, & Thiede, 2005). As described by Rouet and Britt as part of the MD-Trace framework (2011; Britt & Rouet, 2012), the *task model* includes what the goals and subgoals are for reading and writing (e.g., Why are the sources being read? What is the goal for the inquiry task?). In addition to containing the reader's goals, the *task model* subsumes an *activity model* that informs the reader about which specific actions, procedures, or behaviors one might engage in to reach those

goals during the inquiry task. The *activity model* is a representation of how to complete the task (What does completion of the task entail? How should I engage in this task?).

The extent to which readers might develop an *integrated model* or *intertext model* in multiple document contexts will be partially determined by the contents of the *task model,* including the representation of *task goals* and the *activity model*. Although different reading, writing, or processing goals may affect learning in single source contexts, they become even more critical to consider in multiple source contexts. With a single source, the construction of a good-enough *situation model* can often be achieved by representing the original text's intended purpose, structure, or argument in memory. However, when tasked with comprehending an event or phenomenon from multiple sources, the reader must actively impose selection, organization, and transformation to construct a mental model that integrates the information from different texts, rather than more passively creating distinct traces for each individual text (Wiley & Voss, 1996, 1999; Wiley et al., 2009). Often, some or even all of the individual documents a learner has available were written for a purpose other than addressing the questions relevant for the learner's goals. The learner must selectively re-purpose the information in the documents. A reader's *task goals* and *activity model* will determine the extent to which each reader will engage in the process of evaluating the individual texts; selecting, organizing, and transforming relevant information; and re-assembling what is selected into a new coherent model. Thus, the interpretation of the task; and the extent to which readers understand that they need to actively engage in constructive processing, attempt to build connections across ideas, and try to form a coherent, integrated model of the phenomena; are critical determinants of multiple source comprehension (Britt & Rouet, 2012; Wiley & Voss, 1996; 1999).

In general, individuals face several challenges when engaging in multiple source inquiry activities. The primary dependent variables that have been used to assess effective multiple source comprehension include analyses of the quality of the written responses that are generated as part of the inquiry activity, as well as performance on tests of comprehension of the content following an inquiry activity. In addition, measures of sourcing (i.e. consideration of information about when, where, and for what a purpose a source was created), source evaluation (i.e. the ability to discriminate reliable from unreliable sources), or corroboration behaviors are sometimes examined. Research suggests that students are generally unlikely to spontaneously attend to the source of information and note it in their written responses. Students also typically fail to engage in critical evaluation or corroboration of sources. Further, many students tend to engage in superficial representation of the information they read, rather than actively selecting and transforming information to construct a coherent, integrated mental model of the phenomenon. This affects both the quality of their essays and their performance on comprehension tests. These difficulties are the primary issues that instructional manipulations have been designed to address.

Summary of Empirical Research

The main areas of research in this literature have focused on how features of the inquiry task, the task environment, and the instructional context provided prior to engaging in multiple-source inquiry tasks can affect performance.

Inquiry Task Manipulations

Studies that fall under this heading feature manipulations in the reading and/or writing tasks that are assigned to students as the main task for the multiple-document inquiry activity.

Previous work has shown that the writing prompt can alter performance during multiple-document inquiry tasks.

For example, Wiley and Voss (1996; 1999) found that prompting undergraduates to write arguments from information presented in multiple sources about the causes of the Irish Potato Famine (a period of mass starvation, disease, and emigration in Ireland between 1845 – 1852) led to more connected essays, and better comprehension of the material, than when students were prompted to write narratives from the same information presented in a single source. Students prompted to write arguments also wrote essays that contained a higher proportion of transformed sentences, that is, sentences that combined information from the texts in a new way, rather than borrowing sentences directly from the texts, or adding sentences that contained no information from the texts. Wiley (2001) extended these findings, showing that greater benefits were seen when students composed an argument from multiple sources using a two-window browser with additional instructions about why they were being given two windows, versus when they composed a narrative from sources presented in a single window browser. Wiley (2001) and Hemmerich and Wiley (2002) also extended these findings to a document set on a scientific topic and found benefits for writing an argument about the causes of the eruption of Mt. St. Helens in the state of Washington in the United States in 1980 versus writing a narrative, while Wiley et al. (2009) found benefits for writing arguments about the causes of the eruption over writing a description. Le Bigot and Rouet (2007) demonstrated better essay quality from argument than summary writing tasks for a document set on the topic of social influence, and Naumann, Wechsung, and Krems (2009) found better quality essays about the Panamanian Revolution (Panama's separation from Colombia in 1903) in a condition which they referred to as "argumentative," where participants were instructed to "form an opinion about a controversy on

the Panama topic," versus a condition they referred to as "narrative," where participants were instructed to " write a description of the events". Stadtler, Scharrer, Skodzik, and Bromme (2014) found that an argumentation prompt led to more balanced coverage of a medical controversy than a summarization prompt.Their argumentation prompt was "Read the texts attentively to write an argument afterward. The first part of an argument provides a comparative overview of the standpoints and supporting arguments brought forward by the different authors. The second part consists of your own personal view regarding the authors' positions, which should stand on justified grounds." Their summary prompt was "Read the texts attentively to write a summary afterward. A summary is a clearly laid-out overview of the essential contents. Hence, it should inform about what the nine texts are about." Maier and Richter (2016) also found that argumentation prompts led to more balanced reading of sources on either side of a controversy ("build a justified point of view whether or not electromagnetic radiation from cell phones is hazardous") than summarization prompts ("memorize as many facts as possible"). Their argument instruction also included the request to critically evaluate text information to judge the plausibility of arguments.

On the other hand, writing prompts that encourage students to discuss their own opinions in an essay can often lead to poorer comprehension of the materials and lower quality essays. Monte-Sano and De La Paz (2012) had students engage in an inquiry task about the Cold War and found that a situated prompt to write a letter arguing about what is wrong with Communism led to lower quality essays that were less likely to attend to the reconciliation of historical perspectives than prompts that instructed high school students to consider the historical context, compare across documents, or consider the causes of the Cold War. Similarly, Wiley, Steffens, Britt, and Griffin (2014) found both college and high school students wrote better essays about

the Panamanian Revolution when prompted to “write an argument about the factors that contributed to the revolution” than when prompted to “write an argument about the extent to which Teddy Roosevelt and his administration were responsible for the revolution.” Note that while these prompts did not specifically ask students to report their opinions, they may have steered students toward taking a stance on a question of subjective moral evaluation (e.g., “what is wrong with”, who was “responsible for”) rather than a question of fact. Although these studies did not explicitly use the term “opinion” in their prompts, the inherently subjective and personal or ideological nature of the questions used in these studies may have led to poorer essays in terms of addressing the factual matters surrounding the topic.

Stahl, Hynd, Britton, McNish, and Bosquet (1996) explicitly manipulated whether AP high school students received opinion-based reading and writing prompts during an inquiry task about the Vietnam War and the Gulf of Tonkin incident that took place in the waters off the coast of Vietnam in 1964, drawing the United States more directly into the war. They found that students failed to integrate multiple documents or engage in sourcing or corroboration behaviors, regardless of whether they were asked to “write a description” or “write about your opinions regarding” the Gulf of Tonkin incident and resolution. However, students asked to “describe” stuck very close to the information in the texts, and primarily relied on a single text. Students who were asked to write their opinion tended to move away from the original texts, including more generalizations not tied to any single text and more subjective evaluative statements in their writing, and they did so without providing grounding from the texts. In this case, the instruction to “write about your opinions regarding” political actions and policies of governments seemed to steer learners towards their own subjective moral judgments about who was “wrong” or “at fault”.

In some cases, argument-based prompts can lead to poorer outcomes than prompts that encourage students to summarize the information provided in a set of documents. This may be especially problematic when students are fairly naïve to argumentation. When students are given instruction in the skills of argument writing or reasoning from evidence, then clear benefits are seen from argument writing tasks, even among younger students (de la Paz, 2005; Klein & Rose, 2010; Klein & Samuels, 2010; Reisman, 2012). Alternatively, sometimes no differences in learning outcomes have been found due to manipulations in writing prompts with younger students (Strømsø, Bråten, & Britt, 2010), while other times interactions have been seen between task manipulations and individual differences among students in thinking dispositions, reading skills, or prior knowledge.

Griffin, Wiley, Britt, and Salas (2012) found that both individual differences in thinking dispositions and reading skill contributed to middle school students' ability to learn when prompted to write an essay "explaining how and why recent patterns of global temperature are different from what has been observed in the past". In this study, the disposition that was measured was CLEAR thinking (Commitment to Logic, Evidence, and Reasoning) which assessed the extent to which students place value and importance on reasoning about evidence when forming and revising beliefs. Climate change, at least in the United States, is surrounded by political controversy and thus subject to strong personal and emotional biases. A dispositional commitment to evidence-based reasoning could help ensure that a student adopts the correct *task model* of understanding the science surrounding climate change, rather than one of trying to selectively focus on information one can use to defend political opinions on the issue.

Gil, Bråten, Vidal-Abarca and Strømsø (2010a) found that only undergraduates with high prior knowledge about the topic experienced better learning outcomes (as measured by a post-

inquiry inference verification test) from an argument writing prompt that included writing about their opinion (Imagine that you have to write a brief report to other students where you express and justify your personal opinion about how climate changes may influence life on Earth and what are the causes of climate changes) than a summarization prompt (Imagine that you have to write a brief report to other students that summarizes how climate changes may influence life on Earth and what are the causes of climate changes). Otherwise, students who were assigned to write summaries wrote more transformed and integrated essays, and performed better on the inference verification test. Similarly, Bråten and Strømsø (2009) found that only undergraduates with more sophisticated topic-specific epistemologies (readers who considered knowledge about climate change to be tentative) experienced better learning outcomes from an argument-writing task. Gil, Bråten, Vidal-Abarca and Strømsø (2010b) found that epistemologically naïve undergraduates, again using a climate-change-specific epistemology scale, had worse learning outcomes from the opinion writing prompt than the summarization prompt. As in Gil et al. (2010a), students' essays were more transformed and integrated in the summarization condition than the opinion-based argument condition.

Other manipulations have varied whether learners received any intertextual inquiry prompt at all, in comparison to more text-specific prompts. Britt and Sommer (2004) had undergraduates read a document set on the Panamanian Revolution and manipulated whether they responded to "macro-questions" which required attention to relationships across texts to explain what happened and why, or "micro-questions" which required readers to pay attention to specific details within individual texts. Participants who received the macro-questions outperformed those responding to micro-questions when asked to construct a timeline from memory, which the authors employed as a measure of the ability to integrate event information

across multiple documents. Cerdan and Vidal-Abarca (2008) gave undergraduates a document set related to bacteria resistance and found that engaging in an essay task which prompted readers to make intertext connections led to better understanding of the content than short essay questions that prompted readers to consider each text individually. Similarly, Wiley et al. (2014) found that middle school students who completed an intertext timeline activity that required integration of event information from across multiple documents before writing an essay explaining the factors that caused the Panamanian Revolution tended to perform better on an inference verification task than students who did not engage in the timeline activity first.

Taken together, this body of research has demonstrated that, across both science and history domains, tasks which prompt students to generate an argument or consider intertextual connections can foster deeper comprehension from multiple documents than tasks that prompt students to simply describe or summarize what they have read. More specifically, argumentation tasks have led to students including more causes and connections in their essays, more coverage of information across documents in their essays, and more balanced online reading behaviors. Similarly, activities such as engaging in macro-level or intertext tasks while reading can also benefit multiple document comprehension. On the other hand, this research also suggests that argument tasks can be less effective than more descriptive tasks when writing prompts encourage students to discuss their own opinions; or when students are young, have little prior knowledge about the topic, have weak reading skills, or are epistemologically naive.

Manipulations of Features of the Task Environment

Studies that fall under this heading include manipulations that alter the context in which the task is completed including whether the document set contains primary documents, policy arguments, or conflicting information; whether the information is presented in a single chapter

versus in separate documents; and manipulations in presentation format such as whether documents are viewed in a multiple-window environment, or in a fixed order.

Some of the earliest work on multiple source comprehension investigated whether presenting sources as a set of multiple documents, as compared to a single document, would improve comprehension (Wiley & Voss, 1996, 1999). While it might be assumed that processing information from multiple sources would be more demanding, Wiley and Voss found that undergraduates who received information about the Irish Potato Famine in a multiple source format exhibited better understanding of the relations between concepts on verification tasks, and wrote more connected essays, as compared to readers who received the same information as a single textbook-like source. They also wrote essays that contained a higher proportion of transformed sentences, and from this evidence the authors argued that the multiple source format prompted individuals to engage in more constructive processing while writing. In contrast, while writing within a single source format, students tended to rely on the loose and implied temporal connections that already existed in the text and "borrowed" or copied more sentences directly from the texts into their essays. Wiley (2001) extended this work to online sources and found that writing arguments from multiple sources presented within two side-by-side windows encouraged more comparison, corroboration, and integration of ideas across texts versus a single-window browser.

In the context of studies testing the effectiveness of the Sourcer's Apprentice tool, Britt and Aglinskas (2002) also compared single and multiple source presentation. The Sourcer's Apprentice environment provides students with a tutorial in sourcing, corroboration, and contextualization (based on the heuristics described by Wineburg, 1991). It also prompts students to complete notecards for each of the sources in a multiple document set. In the final experiment

in this paper, high school students learned about the Homestead Steel Strike, a serious labor dispute that took place in 1892 near Pittsburgh, Pennsylvania in the United States. Documents were presented either within the Sourcer's Apprentice environment, or participants read the information presented as a single textbook-like document. They found that essays in the Sourcer's Apprentice/Multiple Documents condition were more integrated and were also perceived as being of higher quality when graded by history teachers.

Nokes, Dole, and Hacker (2007) also compared the use of multiple sources to traditional textbooks in the context of a 15-day high school unit on the major events in United States History during the 1920s and 1930s, which covered topics such as the Great Depression, foreign affairs at the time, and the African American movement known as the Harlem Renaissance. In a fully crossed design, students also either received instruction designed to help them learn historical content (content instruction) or instruction designed to help them develop skills in sourcing, corroboration, and contextualization (heuristics instruction). They found that students who read multiple sources performed better on a test of history content than students who read the information in traditional history textbook format, regardless of the type of instruction they received. In addition, all students completed a pre-post multiple document essay task (on either the Battle of Lexington, which was the first military engagement of the United States Revolutionary War in 1775 or the Pullman Strike, which was a national railroad strike in the United States that began in Chicago, Illinois in 1894). Performance on this task revealed that source evaluation and corroboration skills improved most with multiple source heuristics instruction.

In a more recent study, Stadtler, Scharrer, Brummernhenrich, and Bromme (2013) examined the impact of presentation format on the integration of conflicting scientific

information about health risks from high cholesterol levels. Undergraduates were presented with either four separate websites by four authors containing two pairs of opposing views, or were presented with a single website stating only one name as the author. Students in the multiple source condition correctly verified more conflicting points of information than students in the single source condition. Students in the multiple source condition also noted more points of conflict in their essays than students in the single source condition.

In another format manipulation, Naumann et al. (2009) varied whether documents about the Panamanian Revolution had to be read in a fixed order versus the reader being able to select when to read each document. Undergraduates given argumentative writing tasks profited from being able to choose the order of the documents they wished to read, while students given narrative writing tasks benefited from static presentation formats.

The choice of which documents are given to students also matters in several ways. Providing students with a pre-determined document set rather than having them engage in their own Internet search has been shown to increase the amount of time spent on "meaning-making" activities (Cho, 2014). Also, the Internet is rife with unreliable sources on just about every topic. Wiley, Ash, Sanchez, and Jaeger (2011) contrasted a set of pre-selected documents about the causes of volcanic eruptions that contained unreliable sources with a set containing only reliable sources. The latter led to more accurate essays, while the former led to essays that included more misconceptions.

The types of documents in a set can also impact sourcing behaviors. Rouet, Britt, Mason, and Perfetti (1996) manipulated whether a document set about the Panamanian Revolution included primary source documents. Their results revealed that undergraduates who received primary sources were more likely to cite sources in their essays. Paxton (2002) found that

reading an introductory text with a visible author versus an anonymous author led to more sourcing behavior in the essays of AP high school students who wrote about the murder of Julius Caesar from a set of sources.

Document sets can also vary in whether they present opposing viewpoints. Braasch, Rouet, Vibert, and Britt (2012) manipulated whether undergraduates were given brief news reports containing discrepant information or consistent information. They found that students who were given discrepant information used more source information in their essays and demonstrated better source memory compared to students who were given all consistent information. Barzalai and Eshet-Alkalai (2015) manipulated whether undergraduates were exposed to converging or conflicting blog posts about desalination. They found an interaction between the consistency manipulation and the epistemological sophistication of the students. Students who endorsed more sophisticated, evaluative epistemologies (as opposed to less sophisticated absolutist or multiplist epistemologies) were able to write better arguments in the conflicting sources condition.

Other recent work has demonstrated that the presence of policy-related documents can also have an impact on multiple source comprehension. A study by Blaum, Griffin, Wiley, and Britt (2017) manipulated whether middle school students received a policy-related document as part of a document set. Students in both conditions were instructed to “write an essay explaining how and why recent patterns in global temperature are different from the past”. Results revealed that students who received policy documents (related to proposed changes in energy regulations) produced essays that included fewer core concepts and causal connections that addressed the inquiry question than students who did not receive policy documents. A pilot study with high school students showed the same result. These results suggest that opinion-based or policy-based

prompts can be harmful, especially if the topic or the nature of the question steers students toward personal opinions that inherently must go beyond evidence and textual information and encourages reliance upon subjective values or feelings.

Overall, this body of literature indicates that the manner in which the tasks or information is presented is important. For instance, this work demonstrates that presenting scientific or historical information in a multiple-document format can benefit learning and comprehension more than presenting the same information in a single document or textbook-like format. Across these studies, when the information was presented in a multiple-document format, students produced essays that were more integrated, included more connections and transformed sentences, demonstrated better source evaluation and corroboration, and addressed more points of conflict. Additionally, some work has demonstrated that the order in which documents are presented can matter: developing an argument seems to benefit from being allowed to access the documents in any order whereas accessing documents in a fixed order may be more beneficial for developing a thorough description. The type of documents included in the set can also matter. For instance, sets including documents that are less reliable or less relevant for understanding the important causal information underlying a phenomenon can harm comprehension and lead to lower quality essays. The presence of discrepant or inconsistent information can also impact comprehension and essay quality, but this may interact with other variables such as epistemology.

Manipulations of Instructional Context

Studies that fall under this heading include interventions that occurred over several weeks within classrooms. It also includes studies that have used more targeted instructional

manipulations in computer-based tutorials, laboratory experiments, and smaller-scale classroom studies completed within one or two sessions.

Several multi-week instructional interventions have been implemented in high school and middle school history classrooms. For example, De La Paz (2005) tested the benefits of an extended unit on historical reasoning strategies. As part of this unit, middle school students engaged in a mock trial, which was intended to help them understand that there can be varying interpretations of an event. After the mock trial, the teacher introduced historical reasoning strategies that included information about how historians use information within a source to determine its usefulness and trustworthiness, information about comparing the details of one source to another to develop corroboration and notice conflicting views, and information about how to plan and compose argumentative essays. Students in control classes did not receive instruction on either historical reasoning or argument writing strategies, but did read the same social studies texts. After the 12-day unit, students in both conditions completed a multiple document inquiry task on the U.S-Mexican war of 1846-1848 and wrote an opinion essay about whether the United States or Mexico was responsible for instigating the war. Students who received the intervention included more arguments and correct historical information in their essays than students in the control class. A version of this instructional unit was also adapted for use with older students (De La Paz & Felton, 2010). Results from the high school study replicated and extended previous work, showing that students in the intervention condition wrote essays with more elaborated claims and rebuttals, as well as more document citations. De La Paz and colleagues (2017) further explored the effects of a year-long intervention in middle school classrooms and again demonstrated improvement in the argumentation in student essays compared to students who did not receive the intervention.

As mentioned previously, Nokes et al. (2007) conducted a study on another intervention, involving a 15-day history unit on the events and trends of the United States in the 1920s and 30s with high school students. Half of the classrooms received instruction that was focused on content learning while half received instruction focused on heuristics (corroboration, contextualization, and sourcing). Students who received heuristics instruction were more likely to engage in source evaluation and corroboration behaviors on the post-unit multiple-document task than students who received content instruction, especially if they had received the instruction with multiple sources rather than textbooks.

Reisman (2012) studied the effects of a longer term intervention using the Reading-Like-a-Historian (RLH) curriculum. The RLH curriculum focuses on training students how to use document sets to engage in historical inquiry. Students in the RLH condition received between 36–50 Document-Based Lessons over the course of 6 months of instruction in 11th grade history, while the control condition received typical textbook instruction. Students receiving the RLH curriculum outperformed students in the traditional textbook condition in both sourcing and factual knowledge gains. Follow-up analyses further revealed that the RLH intervention was especially beneficial for struggling readers.

Klein and Rose (2010) and Klein and Samuels (2010) studied a year-long intervention for middle school students in science. Half of the students were taught to write arguments or explanations as part of content-area instruction. At the end of the year, all students completed a multiple document inquiry task on plate tectonics, nutrition, or the circulatory system. Students who received argument or explanation instruction throughout the school year learned more from the inquiry tasks than did students in comparison classes. Follow-up analyses suggested that the instructional manipulation most strongly affected argument genre knowledge, which in turn

accounted for variance in science learning.

While these multi-week classroom interventions have been shown to be effective in improving sourcing and content learning, their broad scope, variability in implementation, and inclusion of many different manipulations simultaneously, make it difficult to make attributions about the cause of any improvements. The effects of shorter-term, more targeted instructional manipulations have also been investigated, primarily in the context of computer-based tutorials completed within one or two sessions.

As previously mentioned, the Sourcer's Apprentice environment was designed to support behaviors of sourcing, contextualization, and corroboration by training students on each of these heuristics through short tutorials, and then supporting their application during an inquiry task by requiring students to complete notecards. Britt and Aglinksas (2002) found high school students who used the tool as part of instruction outperformed students who received traditional instruction in sourcing on a transfer task on a new topic. Britt, Wiemer-Hastings, Larson, and Perfetti (2004) found that providing tailored feedback on sourcing (or the lack thereof) in essays written using the Sourcer's Apprentice environment can further increase sourcing behavior in a sample of undergraduates and reduce unsourced usage of original texts.

Building on the work with Sourcer's Apprentice, Wiley et al. (2009) developed a similar unit (SEEK) which included four key behaviors that were found to be important for successful engagement in multiple document inquiry tasks in science. These included attending to the *source* of the information, evaluating the *evidence* that was presented, developing an *explanation* for the phenomena, and integrating new information with prior *knowledge*. In the SEEK training condition, undergraduates completed a practice inquiry task about low carbohydrate diets. They were given information about which aspects of sources to consider when attempting to construct

an explanation as part of an inquiry task, prompted to evaluate the sources, and provided with feedback on their evaluations. They were also presented with the protocol of an example 'peer' who modeled these behaviors. The comparison group completed the same inquiry task on low carbohydrate diets, but were not provided with the SEEK instructional materials. In a second session, both groups completed a second inquiry activity where they were tasked with understanding the causes of the eruption of Mt. St. Helens. Students in the SEEK condition demonstrated better learning from the inquiry activity than students in the comparison condition, included fewer erroneous causes in their essays, and wrote more integrated causal essays. They were also better able to discriminate between reliable and unreliable sources. An alternative version of SEEK training was tested by Graesser et al. (2007), but it was not as effective. The main difference between the two versions was that the ineffective version did not include the example student protocol.

Another instructional study provides further support for the importance of peer examples. Instead of including a single peer model, Braasch, Bråten, Strømsø, Anmarkrud, and Ferguson (2013) used a contrasting-cases approach. Specifically, they presented high school students with two examples of peers engaging in evaluating sources in relation to answering the question whether cell phones pose a health risk. The "better" student protocol included more sophisticated strategies such as checking the information about the source including author information, the venue of the information, and the date of publication. The "poorer" student protocol contained less sophisticated strategies such as checking to see if the title of the article contained the key words they were using. In the experimental condition, students were asked to read through the two example student accounts and compare the strategies used. Control classes received no contrasting case examples or instruction on sourcing strategies. In a second session, students in

both conditions engaged in a multiple document inquiry task on a new topic, the causes of El Nino weather patterns. Students given the contrasting cases instruction included more scientific concepts from more useful sources in their essays, and were also better able to discriminate between more and less reliable documents. The results of this study in combination with the work from Wiley et al. (2009) and Graesser et al. (2007) suggest that providing students with instruction in how to effectively evaluate sources not only improves their ability to detect unreliable information, but can also facilitate the development of mental models of the content. Note that this intervention and the SEEK environment are the two interventions where the content area was science rather than history, and both presented sets of documents that included unreliable sources that presented unscientific information. Interventions in history have not had this explicit feature, but rather tend to use sources that vary more in terms of perspective and involvement with the historical events.

In addition to prompting students to evaluate sources, the SEEK training developed by Wiley et al. (2009) prompted readers to engage in explanation as they read. A series of studies with undergraduates learning from a document set about electricity (Linderholm, Kwon, & Therriault, 2014; Linderholm, Therriault, & Kwon, 2014) manipulated the role of explanation more specifically. In Linderholm, Therriault, and Kwon (2014), students in the control condition were told to "read for comprehension" while students in the experimental condition were told, "As you read the following texts, attempt to explain how circuits work to yourself. Imagine, for example, that you might need to explain the concept of how circuits work to a group of students in a science class." Students who were prompted to engage in explanation while reading did better on a comprehension test. Two follow-up studies by the same authors (Linderholm, Kwon, & Therriault, 2014) tested for the effectiveness of more elaborate instructional conditions. In one

condition, participants received a definition for self-explanation and were urged to use it during reading: “Self-explanation is the process of explaining each sentence of a text using previous text information or your background knowledge to better understand the text. That is, by self-explaining each sentence, you will rephrase it in a way that makes more sense to you or attempt to understand the text by filling in the ‘reason’ for the statement using previous text information or your own background knowledge.” In a second condition, participants received both this definition as well as exposure to the experimenter modeling rephrasing following three sentences from an expository text on another topic (obesity). Again, these conditions were contrasted with a control condition that was simply told to read for comprehension. Both follow-up studies only found benefits of the simpler instructional condition (providing just the definition) over the control condition on test performance. The modeling condition did not result in performance that significantly differed from the control condition. Moreover, both of these studies were only able to show a benefit of the simpler instruction on factual questions, but not on questions that required integration across texts to answer. Thus, the general finding of these studies was that although the instructions did in some cases help to improve memory for the texts, they did not seem to be effective at supporting deeper comprehension or understanding of the topic from a multiple-document inquiry task.

It is possible that the effectiveness of explanation instructions for promoting understanding from multiple documents may have varied as a function of whether readers were allowed to re-read the texts in each study. The follow-up studies (Linderholm, Kwon, & Therriault, 2014) explicitly prevented readers from returning to prior pages in their booklets, whereas this restriction did not appear in the earlier study (Linderholm, Therriault, & Kwon, 2014), nor was it the case in Wiley et al. (2009) where participants had access to all texts

throughout reading and writing.

In addition, providing students with a short definition and examples focusing on re-phrasing or paraphrasing text may not be enough for them to form an appropriate *activity model* of what good explanation entails. The results of another recent study suggest that readers may benefit from a different kind of instruction. Jaeger, Griffin, Britt, and Wiley (2015) developed a pre-reading instruction that informed students that good explanations in science involve considering multiple, linked causes. In their study, all middle school students were given a clear task goal, prompted to use the information from the documents to explain how and why recent patterns in global temperature are different from what has been observed in the past, and encouraged to make connections. However, the half the students who received the pre-reading instruction demonstrated better coverage of and more connectedness among key causal concepts in their essays, as well as better performance on a comprehension test. In contrast with earlier studies by Linderholm and colleagues, this work shows that instruction in explanation-based processing can improve comprehension in a multiple-document context.

These results suggest that even if students understand the general goal of the inquiry task, they may not have a clear understanding of the types of processing that are required to achieve those goals. For example, students might assume that they should seek out the one most probable cause discussed in a particular text rather than find and integrate all the relevant causal factors into a coherent explanation. Importantly, a brief lesson illustrating the multi-causal nature of scientific explanations (using an unrelated topic as the example) yielded notable improvements in learning from multiple documents.

Overall, this area of research demonstrates that the instruction students receive *before* engaging in inquiry tasks is a critical factor to consider for multiple-document inquiry learning

activities. In both extended, classroom-based interventions and shorter-term studies, research has shown that instruction which focuses on developing skills in reasoning and evaluation can foster deeper comprehension and lead to better student essays from multiple-document inquiry tasks. Research in this area has also demonstrated that instructional contexts that prompt students to engage in comparison or explanation, particularly contexts that incorporate examples of desired behaviors, can promote more effective learning from multiple-document contexts by supporting the generation of more appropriate *activity models*.

Summary of Research Findings, Current Challenges, Practical Implications

Research has sought to identify the conditions and interventions that facilitate learning of both content and inquiry skills within multiple-documents environments. Learning can be greater than with traditional single textbook presentations, but not always. There has been increasing attention paid to individual differences, such as in prior knowledge and epistemology, and how these interact with varying features of multiple-document inquiry environments. Although most work has been done with undergraduates, the research suggests that less-sophisticated readers may need support to learn effectively from multiple-document inquiry tasks. Students' perceptions of their task are important for how they engage with multiple sources. Their perceptions are shaped by complex interactions between their *a priori* assumptions and beliefs about comprehension processes (their *activity model*), their perceptions of their *task goals*, and the specifics of any particular multiple-source activity that inform their *task model* (the instructions they are told, the support they are given, the context in which the sources are presented, etc.).

Studies that have manipulated the features of the inquiry task that is given have supported the theoretical importance of the *task model*, and have highlighted that a learner's *task model* can

be sensitive to subtle cues triggered by a single word in the inquiry prompt. Studies that have manipulated features of the task environment, such as changing the type of documents included in a set (e.g., including a policy-related document), have also highlighted how sensitive the *task model* is to subtle changes. Manipulations of the task environment have also shown that the sub-goals and activities a learner engages in (e.g., sourcing and corroboration) can be impacted by how the documents are presented. This suggests that learners may often lack a well-developed *activity model* that would guide them to consciously implement basic sub-goals during inquiry tasks, and instead depend on contextual cues that make different sub-goals salient. Long-term instructional interventions and smaller scale training studies have demonstrated that it is possible to help students develop more appropriate *task goals* and *activity models* that are represented with sufficient enough abstraction to be transferable to new inquiry tasks.

On the whole, learning appears to be greatest when students are instructed to construct and justify what they think is the best explanation of a scientific phenomenon or historical event. However, there are caveats to this. If tasked with constructing explanations rather than narratives, then students appear to benefit from presentation formats that allow them some control over how they navigate between the documents. Also, the exact wording of the prompt is critical. Opinion prompts can be harmful, especially if the topic or the nature of the question steers students toward personal opinions that inherently must go beyond evidence and textual information, and encourages students to rely upon their subjective values or feelings. Even prompts that mention “argument” may lead to similar problems in such contexts. In science, this means being clear about whether the task is one of scientific understanding or policy advocacy. Presenting an opinion or argument about the causes or effects of climate change is fundamentally different from presenting an opinion or argument about what society should do to deal with the

issue. Features of the inquiry question or documents could steer students toward a policy-focus at the expense of comprehending scientific phenomena. Policy opinions are a type of attitude preference that have an inherent basis in affective and personal morals that can produce different reactions to texts than knowledge-based beliefs about factual matters (Wolfe & Griffin, in press). For example, the oft misinterpreted *belief-polarization* effect actually showed that people's policy attitudes about laws allowing capital punishment became more extreme in the opposite direction of the attitude-conflicting texts they read, but their beliefs on the factual matter of crime deterrence became less extreme and changed in the direction of belief-conflicting texts (Lord, Ross, & Lepper, 1979). Thus, inquiry tasks need to take care in considering whether the prompt and the texts relate to preferences or opinions, versus matters of fact.

In history, the opportunity for confusion may be even greater, because the phenomenon to be understood entails people acting in ways that impact others, which often triggers moral concerns that blur the line between understanding causes of events and assigning ethical blame for their negative consequences. There is a critical distinction between what the causes of a war were and whether it should have occurred. If students are to be encouraged to "take a side" as part of any inquiry task, it needs to be made clear to them what exactly they are taking a side about. Nevertheless, the "is/ought" or "factual/ethical" distinction might manifest itself differently in science and history.

The fact that science-related interventions tend to be more focused on training students to selectively rely on information from trustworthy sources raises another issue. In history, primary sources are critical for developing an understanding of the events related to the phenomena to be understood. However, the nature of primary sources is that they represent viewpoints from individuals who may have played different roles or had access to different information related to

the target phenomenon. Primary historical sources from authors with biased interests are often the most important for developing an understanding of the events and the motives of those who caused them. Primary source authors sometimes lay ethical blame on other parties, yet this can reveal those authors' own motivations that are critical to understanding the causes of their actions that produced some focal event. In contrast, sources in which authors display strong personal, ethical, or ideological biases on scientific questions are often viewed as untrustworthy and as a source of misconceptions that are to be avoided in favor of more objective and reliable sources. For example, authors of original articles reporting the results of scientific research are not expected to have or discuss their personal stake in evidence they are presenting. As such, sourcing may play different roles in science and history, which implies that instruction in sourcing skills may need to be tailored to the different types of sources that are used in different disciplines.

One beneficial aspect of multiple-document inquiry tasks is that they provide an opportunity for students to engage in active or constructive comprehension processes. With only a single source, the mental representation of the text can be achieved fairly passively by representing the original text's intended purpose, structure, or argument in memory. However, when tasked with comprehending an event or phenomenon from multiple documents, the reader must actively impose selection, organization, and transformation to construct a representation that integrates the information from different texts. To answer an inquiry question, the learner must also selectively re-purpose the information from the original documents. For many studies, the primary evidence for the reader having engaged in constructive processing during the inquiry activity comes from detailed analysis of the inquiry product. Although trace methods such as think-aloud protocols, navigation logs, or scanpaths are generally thought of as measures that can

reflect the online processing of information, a variety of measures of essay quality have been developed with the goal of assessing the extent to which students attempted to integrate and transform information as they wrote. One measure has been the incidence of connections and connectives included in each essay (Britt & Aglinskas, 2002; Voss & Wiley, 1997; Wiley & Voss, 1999). This serves as an index of the extent to which students attempted to connect or integrate ideas that were previously unconnected in the original texts. Students who demonstrate better understanding of the material on comprehension tests tend to write essays that have more connected ideas and more causal connections (Voss & Wiley, 1997, 2000; Wiley, 2001; Wiley & Voss, 1999).

Another measure of constructive activity considers the origin of information included in the essay, and the extent to which students plagiarize or copy information from the original texts. In one approach, each sentence is scored as to whether or not it contains a connection between idea units that were presented separately in the reading materials. The content of each sentence is classified into one of three categories: transformed, added, or borrowed (Wiley & Voss, 1996, 1999). Sentences that combine some presented information with a new claim or fact, or that integrate two bits of presented information that were not previously connected, are classified as transformed. A sentence is coded as added when it contains only novel information. Sentences that are taken directly from, or are paraphrased from, the original material are classified as borrowed. Students who demonstrate better understanding of the material on comprehension tests tend to write essays that contain a lower proportion of borrowed or copied sentences (Voss & Wiley, 1997, 2000; Wiley, 2001; Wiley & Voss, 1999). Similarly, using automated plagiarism detection techniques, Britt et al. (2004) identified the use of unsourced copied material and excessive quoting as two primary deficiencies when students compose essays from multiple

sources. Also using another automated approach, the findings of Foltz, Britt, and Perfetti (1996) suggested that similarity to an expert model is more important for essay quality than similarity to the original texts. Finally, Wiley et al. (2017) found a significant negative correlation between plagiarism scores and explanation quality. This suggests that the quality of the inquiry product suffers as students fail to transform information and fail to engage in constructive activity as they write from multiple documents.

Constructive activity during inquiry is also reflected by the extent to which the student essay is responsive to the inquiry prompt and re-purposes text from the documents, rather than using segments of text as written for their original purpose. When the documents do not contain verbatim phrases that can be used to directly address the inquiry prompt, students who merely copy text ideas without translation or integration will tend to write essays that focus on the purpose of the original documents rather than the inquiry prompt. In Wiley et al. (2017), whether student essays addressed the inquiry prompt predicted the quality of essays as well as learning outcomes. In fact, in that study the document set was deliberately designed so that each document was originally written for a different purpose that could not be used to directly address the inquiry prompt. Many studies have designed their document sets and inquiry prompts in such a way, as this is a useful method that helps the researcher to distinguish between essays that merely copy versus those that re-purpose and integrate the information.

Regardless of whether online trace measures are collected, or whether measures of constructive processing are derived from an analysis of inquiry products such as essays, a key for future research is that studies endeavor to include both measures of processing and of learning outcomes so that it can be better understood what particular behaviors lead to success, and which activities need to be encouraged or supported to promote effective learning from multiple-

document inquiry tasks. An additional important direction for future research is to continue to integrate work using experimental manipulations, processing measures, and assessments of learning outcomes, with the exploration of individual differences. Several findings suggest that individual differences in epistemic dispositions may affect learning from multiple-document inquiry tasks by prompting different students to have different *task models* (Bråten et al., 2011; Griffin et al., 2012). Readers who are disposed to using evidence and reasoning to update their beliefs may adopt a very different *task model* from those who do not. Similarly, students with more sophisticated epistemic beliefs may view a multiple-document inquiry task as an exercise in corroboration, seeking coherence, and looking for evidence to support claims, whereas students with less sophisticated epistemic beliefs may see the goal of an inquiry task as simply finding the "right" answer verbatim within the documents. A final interesting question for future research is discovering the conditions under which engaging in multiple-document inquiry activities might serve to alter students' epistemic beliefs and thinking dispositions.

Multiple-document inquiry tasks provide an opportunity to help students not only gain a richer understanding of content, but also to develop skills that are important for seeking, selecting, evaluating, re-purposing, and integrating information from various sources. Such skills are vital to life-long learning outside of classrooms. However, students need help developing these skills and, if simply thrown into a multiple-document environment without proper scaffolds, may learn even less content than from traditional textbook- and lecture-based instruction. Researchers and educators must carefully consider the nature of the documents provided; the training of behaviors like sourcing, corroborating, constructing arguments, and explaining; and the wording of the inquiry question so that it may prompt constructive behaviors like integrating information in order to develop an argument or explanation, without inviting

subjective personal opinions on non-factual issues. Without these considerations, students may fail to reap the benefits of learning from multiple-document inquiry activities.

References

Barzilai, S., & Eshet-Alkalai, Y. (2015). The role of epistemic perspectives in comprehension of multiple author viewpoints. *Learning and Instruction*, *36*, 86-103.

Blaum, D., Griffin, T. D., Wiley, J., & Britt, M. A. (2017). Thinking about global warming: Effect of policy-related documents and prompts on learning about causes of climate change. *Discourse Processes, 54,* 303-316.

Braasch, J. L. G., Bråten, I., Strømsø, H. I., Anmarkrud, Ø., & Ferguson, L. E. (2013). Promoting secondary school students' evaluation of source features of multiple documents. *Contemporary Educational Psychology*, *38*, 180-195.

Braasch, J. L. G., Rouet, J. F., Vibert, N., & Britt, M. A. (2012). Readers' use of source information in text comprehension. *Memory & Cognition*, *40*, 450-465.

Bråten, I., & Braasch, J. L. G. (2018/this volume). The role of conflict in multiple source use. In J. L. G. Braasch, I. Bråten, & M. T. McCrudden (Eds.), *Handbook of multiple source use*. New York: Routledge.

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

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*, 1-31.

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

Britt, M. A., & Rouet, J.-F. (2012). Learning with multiple documents: Component skills and their acquisition. In M. J. Lawson & J. R. Kirby (Eds.), *The quality of learning:*

Dispositions, instruction, and mental structures (pp. 276–314). New York, NY: Cambridge University Press.

Britt, M. A., & Sommer, J. (2004). Facilitating textual integration with macro-structure focusing tasks. *Reading Psychology, 25,* 313–339.

Britt, M. A., Wiemer-Hasting, P., Larson, A., & Perfetti, C. A. (2004). Automated feedback on source citation in essay writing. *International Journal of Artificial Intelligence in Education, 14*, 359–374.

Cerdán, R., & Vidal-Abarca, E. (2008). The effects of tasks on integrating information from multiple documents. *Journal of Educational Psychology*, *100*, 209-222.

Cho, B. Y. (2014). Competent adolescent readers' use of internet reading strategies: A think-aloud study. *Cognition and Instruction*, *32*, 253-289.

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

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

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,* 31-52.

Foltz, P. W., Britt, M. A., & Perfetti, C. A. (1996) Reasoning from multiple texts: An automatic analysis of readers' situation models. In G. W. Cottrell (Ed.) *Proceedings of the 18th*

annual conference of the Cognitive Science Society (pp. 110-115). Mahwah, NJ: Erlbaum.

Gil, L., Bråten, I., Vidal-Abarca, E., & Strømsø, H. I. (2010a). Summary versus argument tasks when working with multiple documents: Which is better for whom? *Contemporary Educational Psychology*, *35*, 157-173.

Gil, L., Bråten, I., Vidal-Abarca, E., & Strømsø, H. I. (2010b). Understanding and integrating multiple science texts: Summary tasks are sometimes better than argument tasks. *Reading Psychology*, *31*, 30-68.

Graesser, A. C., Singer, M., & Trabasso, T. (1994). Constructing inferences during narrative text comprehension. *Psychological Review, 101,* 371–395.

Graesser, A. C., Wiley, J., Goldman, S. R., O'Reilly, T., Jeon, M., & McDaniel, B. (2007). SEEK Web tutor: Fostering a critical stance while exploring the causes of volcanic eruption. *Metacognition and Learning*, *2*, 89-105.

Griffin, T.D., Wiley, J., Britt, M.A., & Salas, C. (2012). The role of CLEAR thinking in learning science from multiple-document inquiry tasks. *International Electronic Journal of Elementary Education*. *5*, 63-78.

Hemmerich, J., & Wiley, J. (2002). Do argumentation tasks promote conceptual change about volcanoes? *Proceedings of the 24th annual conference of the Cognitive Science Society* (pp. 453-458). Hillsdale, NJ: Erlbaum.

Jaeger, A. J., Griffin, T. D., Britt, M. A., & Wiley, J. (2015, July). *Making connections: Improving student learning about climate change*. Paper presented at the 25th Annual Meeting of the Society for Text & Discourse, Minneapolis, MN.

Kintsch, W. (1998). *Comprehension: A paradigm for cognition*. Cambridge: Cambridge University Press.

Klein, P. D., & Rose, M. A. (2010). Teaching argument and explanation to prepare junior students for writing to learn. *Reading Research Quarterly, 45*, 433-461.

Klein, P. D., & Samuels, B. (2010). Learning about plate tectonics through argument-writing. *Alberta Journal of Educational Research, 56,* 196-217.

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

Linderholm, T., Kwon, H., & Therriault, D. J. (2014). Instructions that enhance multiple-text comprehension for college readers. *Journal of College Reading and Learning, 45*, 3-19.

Linderholm, T., Therriault, D. J., & Kwon, H. (2014). Multiple science text processing: Building comprehension skills for college student readers. *Reading Psychology, 35*, 332-356.

Lord, C. G., Ross, L., & Lepper, M. R. (1979). Biased assimilation and attitude polarization: The effects of prior theories on subsequently considered evidence. *Journal of Personality and Social Psychology, 37*, 2098-2109.

Maier, J., & Richter, T. (2016). Effects of text-belief consistency and reading task on the strategic validation of multiple texts. *European Journal of Psychology of Education, 31,* 479-497.

Monte-Sano, C., & De La Paz, S. (2012). Using writing tasks to elicit adolescents' historical reasoning. *Journal of Literacy Research, 44*, 273-299.

Naumann, A. B., Wechsung, I., & Krems, J. F. (2009). How to support learning from multiple hypertext sources. *Behavior Research Methods, 41,* 639–646.

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*, 492-504.

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

Perfetti, C. A., Rouet, J.-F., & Britt, M. A. (1999). Towards a theory of documents representation. In H. van Oostendorp & S. R. Goldman (Eds.), *The construction of mental representations during reading* (pp. 99–122). Mahwah, NJ: Erlbaum.

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

Rouet, J.-F., & Britt, M. A. (2011). Relevance processes in multiple document comprehension. In M.T. McCrudden, J.P. Magliano, & G. Schraw (Eds.), *Text relevance and learning from text* (pp 19–52). Greenwich, CT: Information Age.

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,* 478–493.

Stadtler, M., Scharrer, L., Brummernhenrich, B., & Bromme, R. (2013). Dealing with uncertainty: Readers' memory for and use of conflicting information from science texts as function of presentation format and source expertise. *Cognition and Instruction, 31,* 130-150.

Stadtler, M., Scharrer, L., Skodzik, T., & Bromme, R. (2014). Comprehending multiple documents on scientific controversies: Effects of reading goals and signaling rhetorical relationships. *Discourse Processes, 51*, 93-116.

Stahl, S. A., Hynd, C. R., Britton, B. K., McNish, M. M., & Bosquet, D. (1996). What happens when students read multiple source documents in history? *Reading Research Quarterly*, *31*, 430-456.

Strømsø, H. I., Bråten, I., & Britt, M. A. (2010). Reading multiple texts about climate change: The relationship between memory for sources and text comprehension. *Learning and Instruction*, *20*, 192-204.

Voss, J. F., & Wiley, J. (1997). Developing understanding while writing essays in history. *International Journal of Educational Research, 27,* 255-265.

Voss, J. F., & Wiley, J. (2000). A case study of developing historical understanding via instruction: The importance of integrating text components and constructing arguments. In P. Stearns, S. Wineburg, & P. Seixas (Eds.), *Knowing, teaching and learning in History* (pp. 375-389). New York, NY: NYU Press.

Wiley, J. (2001). Supporting understanding through task and browser design. In *Proceedings of the 23rd annual conference of the Cognitive Science Society* (pp. 1164-1169). Hillsdale, NJ: Erlbaum.

Wiley, J., & Ash, I. (2005). Multimedia learning in history. In R. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 375-391). New York, NY: Cambridge University Press.

Wiley, J., Ash, I.K., Sanchez, C.A., & Jaeger, A. (2011). Clarifying readers goals for learning from expository science texts. In M. McCrudden, J. Magliano, & G. Schraw (Eds.), *Text relevance and learning from text* (pp. 353-374). Greenwich, CT: Information Age.

Wiley, J., Goldman, S. R., Graesser, A. C., Sanchez, C. A., Ash, I. K., & Hemmerich, J. A. (2009). Source evaluation, comprehension, and learning in Internet science inquiry tasks. *American Educational Research Journal*, *46*, 1060-1106.

Wiley, J., Griffin, T. D., & Thiede, K. W. (2005). Putting the comprehension in metacomprehension. *Journal of General Psychology*, *132,* 408-428.

Wiley, J., Hastings, P., Blaum, D., Jaeger, A. J., Hughes, S., Wallace, P., Griffin, T. D., & Britt, M. A. (2017). Different approaches to assessing the quality of explanations following a multiple-document inquiry activity in science. *International Journal of Artificial Intelligence in Education*. Advance online publication. doi: 10.1007/s40593-017-0138-z.

Wiley, J., Steffens, B., Britt, M. A., & Griffin, T. D. (2014). Writing to learn from multiple-source inquiry activities in history. In G. Rijlaarsdam (Series Ed.), P. Klein, P. Boscolo, C. Gelati, & L. Kilpatrick (Volume Eds.) (Eds.), *Studies in writing, writing as a learning activity* (pp. 120-148). Leiden, Netherlands: Emerald Group Publishing Limited.

Wiley, J., & Voss, J. F. (1996). The effects of "playing" historian on learning in history. *Applied Cognitive Psychology*, *10*, 63-72.

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

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.

Wolfe, M. B., & Griffin, T. D. (in press). Beliefs and discourse processing. To appear in M. F. Schober, D. N. Rapp, & M. A. Britt (Eds.), *Handbook of discourse processes* (2nd ed.), New York, NY: Routledge.