INTERNATIONAL LITERACY ASSOCIATION READING RESEARCH QUARTERLY

Evidence Use in Argument Writing Based on Multiple Texts

Hongcui Du

Alexandra List

The Pennsylvania State University, State College, USA

ABSTRACT

In this study, we focused on one particular challenge in argumentative writing based on multiple texts: evidence use. Specifically, we first examined undergraduate students' strategy reports when processing evidence presented in conflicting and comparably trustworthy texts. Both surface-level (i.e., restating) and deep-level (e.g., evaluation, elaboration) evidence-related processing strategies were identified in students' typed strategy reports. Second, we examined students' use of text-based evidence in their writing. Students were found to have limitations in both the quantity and quality of evidence use in writing. Finally, we associated students' evidence-related processing strategies with their evidence use in writing. Only the quality of students' evidence use in writing was significantly predicted by their evidence-related processing strategies. However, significant predictors of the quality of evidence use in writing included both surface- and deep-level strategies. Implications for instruction are discussed.

Reading–writing tasks, or assignments that ask students to compose written responses based on multiple texts, are commonly assigned as a way of engaging students in synthesis writing, considered to be a critical competency for learning in the 21st century (Bråten, Strømsø, & Britt, 2009; Cerdán & Vidal-Abarca, 2008; Wiley & Voss, 1999). In part, reading–writing tasks require that students select and connect key information across texts and communicate this information to an external audience (Applebee, 1984; Klein, 1999; Spivey, 1990). Despite its importance, students have been found to struggle with writing effectively based on multiple texts. Challenges include students' difficulties with communicating complex argument structures (e.g., presenting both sides of an argument; Anmarkrud, Bråten, & Strømsø, 2014) and with integrating multiple texts in the responses that they compose, rather than describing texts in a sequential fashion (List, Du, Wang, & Lee, 2019). One particular challenge identified in prior work is students' difficulties with using and describing evidence presented across texts, in their writing (De La Paz, Ferretti, Wissinger, Yee, & MacArthur, 2012; Du & List, 2020; Monte-Sano & De La Paz, 2012). This challenge was the focus of the present study. In particular, building on prior work, we used students' written responses and typed strategy reports to examine the nature of evidence use during processing and in the written responses that students composed. As such, we had three primary goals. First, we aimed to identify students' processing strategies when asked to complete a multiple-text reading–writing task. Second, we were interested in students' actual evidence use in writing. Finally, we sought to

INTERNATIONAL LITERACY ASSOCIATION

Reading Research Quarterly, 56(4)
pp. 715–735 | doi:10.1002/rrq.366

associate students' processing strategies with their evidence use in writing.

Reading—Writing Tasks

Reading–writing can be defined as a hybrid activity during which a person reads textual sources with the intention of producing a text of their own (Flower et al., 1990; Spivey, 1990). More than the sum of their parts, reading–writing tasks encompass the dynamic interactions between reading and writing, two highly interdependent and iterative processes. Spivey (1990) suggested that when completing reading–writing tasks, students engage in three core processes—selecting, organizing, and connecting information—both when reading textual sources and when composing a written response. When reading, students determine which information in texts is important or relevant for task satisfaction, organize selected information in the cognitive representations that they construct, and engage prior knowledge to help them better understand text content (i.e., integrate; Afflerbach & Cho, 2009; Cho, Afflerbach, & Han, 2018). In turn, when writing, students determine which information from source text(s) they want to include in their written products, organize selected information in their written responses, and describe textual information, as connected with prior knowledge (Monte-Sano & De La Paz, 2012). This final, connecting step may also involve the linking or integration of information across texts and the description of such integration through writing.

Likely due to the complexity associated with reading–writing task completion, students have been found to struggle when asked to complete such tasks. This may be the case for a variety of reasons. For one, students may be challenged by the informational demands of searching for, processing, and comprehending multiple texts (Gil, Bråten, Vidal-Abarca, & Strømsø, 2010). For another, writing may pose a variety of conceptual (e.g., formulating an argument) and compositional (e.g., choosing the right word) demands on learners (Anmarkrud et al., 2014; Anmarkrud, McCrudden, Bråten, & Strømsø, 2013; Mateos et al., 2020). Nevertheless, in this article, we focus on a unique challenge associated with completing reading–writing tasks: transforming text-based evidence into writing.

Challenges in Reading—Writing Tasks

Several aspects of transforming text-based information into writing may pose a challenge for students. For one, they need to be able to comprehend a set of texts provided and determine which information in texts is most relevant or important for response composition. Whereas students have been found to effectively distinguish relevant from irrelevant information during processing and to deploy different strategies accordingly (Anmarkrud et al., 2013; Bråten, McCrudden, Stang Lund, Brante, & Strømsø, 2018; Cerdán & Vidal-Abarca, 2008), few studies have directly examined which information in texts students ultimately choose to include in the written responses that they compose.

For another, students need to be able to paraphrase selected information, translating it into language of their own. In a study examining undergraduates' note-taking and comprehension performance when reading multiple texts, Hagen, Braasch, and Bråten (2014) identified two levels of transformation in students' notes. Specifically, at the first level (i.e., paraphrase), students' notes reflected a verbatim repetition of text information, changing only a few words without changing their meaning. At the second level (i.e., elaboration), two types of transformations were identified: intratextual elaborations (i.e., connecting information from the same text, connecting text-based information with prior knowledge) and intertextual elaborations (i.e., connecting information across different texts). Hagen et al. found students' notes to mostly be paraphrased (mean [*M*] = 5.45), with relatively few intratextual (*M* = 3.75) or intertextual (*M* = 0.36) elaborations evidenced, demonstrating little to no transformation of text-based information. This indicates that students experience difficulties with translating authors' content into their own, original writing.

Moreover, even after text-based information is selected and rephrased, such information needs to be organized and reorganized in relation to task demands, the thesis or claim that students develop, and other evidence presented across texts. Britt and Sommer (2004) systematically manipulated texts so students needed to switch between two texts 15 times to develop a conceptually coherent story of the building of the Panama Canal in accurate, chronological order. However, students' free recall of text-based information was found to have only a limited number of switches, or transitions between texts, with students including only an average of 3.40 switches, out of a possible 15, when presented with a single integrative text, and students including only 2.28 switches when required to iterate between two texts. This demonstrates the difficulties that students experience with organizing text-based information, even when explicitly instructed to do so and when this information is presented in only one text, rather than across multiple texts.

In addition, during the organization and reorganization of text-based information, the integration of information across texts may also be required. *Integration* refers to students' explicit connecting of similar, different, or otherwise related content across texts and its attribution to particular sources of origin (i.e., Britt, Perfetti, Sandak, & Rouet, 1999; List & Alexander, 2017; Perfetti,

Rouet, & Britt, 1999). Integration has been found to be a challenge for undergraduates when reading multiple texts (Gil et al., 2010), with students both identifying a relatively limited number of connections across texts (List, Du, & Lee, 2020; List et al., 2019) and often failing to accurately track information back to its source of origin (Braasch, Rouet, Vibert, & Britt, 2012; List, Alexander, & Stephens, 2017).

Finally, students may also need to evaluate the information provided in texts to determine whether it is suitable for use in the written responses that they compose. Nevertheless, undergraduates have been found to evaluate texts and their content rarely during multiple-text use (Bråten et al., 2009; Kammerer & Gerjets, 2012; List et al., 2017). Moreover, limited work has explicitly considered whether students' use of processing strategies while evaluating information during reading is associated with the information that they ultimately include in the responses that they compose. In this study, we focused on students' selection, organization, connection, and evaluation of one particular type of text-based information during reading and writing: evidence. We define evidence as information about individuals or the world, gathered in some systematic or intentional way, that can be used to support a claim (Toulmin, 1958) or to inform readers. We examined how students processed text-based evidence during reading and translated this evidence to their writing. In this way, we also associate evidence-related processing during reading with evidence use in writing.

Evidence Use in Writing

Although students' transfer and transformation of evidence from the texts that they read into their own writing has received comparatively little attention in the literature, evidence use has been a prominent dimension along which students' written responses have been scored. Most commonly, students' written products have been rated only according to whether they included text-based evidence or not (e.g., undergraduates: Anmarkrud et al., 2014; elementary students: Reznitskaya, Kuo, Glina, & Anderson, 2009). Indeed, such circumspect considerations of evidence, as only present or absent, have commonly been included in general coding schemes examining the quality of undergraduates' argumentative writing (e.g., Strømsø & Bråten, 2014; Strømsø, Bråten, Britt, & Ferguson, 2013).

An exception to this is work by Wiley and Voss (1999), who specifically examined the textual origin of information included in undergraduate students' written responses. In particular, Wiley and Voss coded each sentence in students' writing as borrowed (i.e., when directly taken from texts or paraphrased), added (i.e., when composed based on prior knowledge), or transformed (i.e., when connecting text-based information with prior knowledge or with other pieces of information in a novel way), according to its origin within a set of multiple texts. Most sentences in students' writing were found to be borrowed, rather than transformed, pointing to the difficulties that students experience with reformulating and integrating text-based information. Although Wiley and Voss examined information origin generally, rather than looking at evidence more specifically, their study points to some of the difficulties that students may experience with text-based evidence use during writing.

Beyond Wiley and Voss's (1999) study, other work has sought to specifically evaluate the quality of students' evidence use in writing. Monte-Sano (2010) examined adolescents' evidence use in historical writing. Specifically, she used students' written responses to derive five dimensions according to which evidence use could be evaluated: accuracy (i.e., correctness of interpretation), persuasiveness (i.e., convincingness or evidence strength), sourcing (i.e., reference to the text of origin), corroboration (i.e., integration of evidence across texts), and contextualization (i.e., placement of evidence into a historical context). Nevertheless, although she provided a potential coding scheme for the evaluation of evidence use, Monte-Sano did not score students' written responses based on this coding scheme, per se.

Building on Monte-Sano's (2010) work, De La Paz et al. (2012) scored students' essays composed based on multiple texts according to the quality of their evidence use. Using a 6-point rubric, the researchers assigned points for explicitly referencing a document or an author, referencing a document or an author correctly or in a relevant fashion, using evidence to validate a claim, using evidence in a balanced fashion (i.e., providing evidence for and against an argument), and evaluating evidence quality. Examining a sample of eighth- and 11th-grade history students, De La Paz et al. found that students, across grade levels, struggled with providing balanced evidence and with evaluating evidence quality. At the same time, better writers were found to use more evidence in writing as compared with weaker writers. Nevertheless, there remains the open question of what factors might be associated with students' evidence use in writing.

In an intervention study, Monte-Sano and De La Paz (2012) found that prompting students to engage in different processing (i.e., sourcing, corroboration, causal analysis) improved their historical writing and reasoning. Nevertheless, the researchers examined students' processing of historical documents as holistic pieces of historical evidence, rather than targeting students' reasoning about evidence embedded within text. Moreover, Monte-Sano and De La Paz cued students' strategic processing through targeted writing prompts, leaving open the question of which strategies students may spontaneously engage when asked to reason about evidence in multiple texts,

and the extent to which such strategy use is associated with writing performance.

In this study, we built on these prior studies to evaluate students' evidence use in writing. Specifically, we examined both the quantity (i.e., total amount of evidence provided in writing) and quality (i.e., use of evidence that was commonly cited but differentially interpreted across source texts) of students' evidence use in writing. Moreover, in addition to examining students' written products, we further examined students' evidence-related processing while reading, via a typed strategy elicitation protocol. In doing so, we aimed to understand what strategies students use when reasoning about evidence provided in texts. Finally, we associated the quantity and quality of students' evidence use in writing with their preceding evidence-related reasoning and evaluation during reading. Examining such associations helped us better understand the processes underlying students' evidence-based writing and identify potential targets for instruction and intervention.

Processing Strategies for Evidence Evaluation

Whereas a number of studies have examined students' strategic processing during multiple-text use (e.g., Bråten & Strømsø, 2003, 2011; Wineburg, 1991), less has been done to specifically examine students' attendance to, reasoning about, and evaluation of text-based evidence during reading. For the most part, students' evaluations of text content, including text-based evidence, have been found to be general judgments, often distinguishable from students' evaluations of source or author credibility (Bråten et al., 2009; Rouet, Britt, Mason, & Perfetti, 1996). Nevertheless, some studies have examined the particular criteria that students use when evaluating specific content or evidence in texts.

For instance, Walraven, Brand-Gruwel, and Boshuizen (2010) examined secondary school students' evaluation of information when researching a health topic and a history topic on the internet. Specifically, the researchers developed a list of criteria that students could use when evaluating information online: information usability (i.e., language used, task relevance, intended audience, recency, information volume), verifiability (i.e., author information, references, relation to other sources, prior knowledge), and reliability (i.e., its factual basis, objectivity, authors' goals). Based on students' think-aloud utterances while searching online, Walraven et al. found that without intervention, students mostly evaluated content based on its task relevance, reflected in 74.73% of the evaluations produced. In contrast, students' evaluations were rarely based on the objectivity of information (2.20% of evaluations) or its ability to be corroborated across sources (3.30%).

Similarly, Goldman, Braasch, Wiley, Graesser, and Brodowinska (2012) identified six types of evaluations in students' think-aloud utterances, generated during a multiple-document inquiry task: information relevance (i.e., whether information was pertinent to the topic of the task), information reliability (i.e., whether information was scientifically sound), source credibility (i.e., whether information came from a credible author), information or text quality (i.e., text content, such as the provision of evidence or examples), students' affective reactions (i.e., emotional responses to information), and search outcomes (i.e., whether students' needs were satisfied by the information found). Goldman et al. found that students evaluated texts primarily based on information relevance (33%) and text quality (31%), with only 6% of evaluations considering source credibility.

These previous studies indicated that although students can evaluate information according to a number of criteria, including its trustworthiness, quality, and factual basis, they rarely do so, attending primarily to information relevance during task completion. Moreover, these studies did little to consider what other types of strategies students may engage when reasoning about text-based evidence. In this study, we focused more broadly, beyond investigations of evidence evaluation, to consider other evidence-based reasoning strategies and their relation to students' writing performance. In doing so, we sought to identify both strategies that are associated with evidence evaluation and those that specifically contribute to students' evidence use in writing. Indeed, we considered both students' critical evaluation of evidence and effective communication about evidence through writing to be key academic outcomes to target through investigation, instruction, and intervention.

The Present Study

Drawing on work examining evaluative processing during reading and identifying evidence use as an important feature of students' writing, we had three primary goals in this study. First, we examined students' evidence-related strategy use during reading, including their attention to and evaluation of text-based evidence. Second, we analyzed both the quantity and quality of students' evidence use in their writing. Specifically, we were interested in the volume and accuracy of evidence included in students' written responses and how such evidence was used to support arguments or to rebut counterarguments. Finally, we associated students' evidence-related strategy use during reading with their evidence use in writing.

Although a plurality of definitions exists (e.g., Hoeken & Hustinx, 2003; Sandoval & Millwood, 2005), evidence has most commonly been defined in relation to a claim, serving as its justification (Toulmin, 1958). For the purposes of this study, we define evidence as information

about individuals or the world, purposefully gathered in some way, which, although varying widely in quality and modality, can be independently generated or verified by some other entity. As such, we expand on Toulmin's (1958) definition to recognize that although evidence may be provided in support of a claim, it may also be used in expository texts more generally to inform, rather than persuade, readers. Therefore, we define evidence not according to its relation to a claim but rather in terms of its constitution, as including deliberately gathered information about the world. Moreover, we recognize that evidence includes both information that may be directly generated by learners themselves (e.g., through data collection) or provided to students secondhand (e.g., through a text or a demonstration). In this study, we were interested in students' attention to and use of secondhand evidence appearing in texts and in students' interpretation and evaluation of such evidence. Three research questions guided this study:

1. What are students' evidence-related processing strategies, as captured by a typed strategy elicitation protocol, when reading multiple texts?
2. What characterizes students' evidence use in writing?
3. What is the association between students' evidence-related processing strategies, reported during reading, and the quantity and quality of their evidence use in writing?

Method

Participants

Participants were 78 undergraduate students from a large research university in the Northeastern United States. Undergraduate students, in particular, were targeted for this study because argumentative writing and writing based on empirical evidence represent key features of college-level academic writing, across domains (Anmarkrud et al., 2014; Barzilai, Tzadok, & Eshet-Alkalai, 2015). Although students' general writing performance was not assessed as a part of this study, we expected all students to be at least somewhat competent college-level writers, as evidenced by their successful admission to a selective university, which required them to complete the Evidence-Based Reading and Writing portions of the SAT. For this university's 2017–2018 class, corresponding to the majority of participants in this sample, admitted students in the 25th percentile scored 580, out of a possible 800 points, on the Evidence-Based Reading and Writing portion of the SAT, whereas admitted students in the 75th percentile scored 660. According to the College Board (n.d.), the SAT benchmark for college readiness is a score of 480, suggesting that students in our sample were prepared to engage in evidence-based reading–writing at the college level.

Students were, on average, 18.74 years old (standard deviation [*SD*] = 1.06 years). The majority of participants were female (93.24%, $n = 69$), freshmen (79.73%, $n = 59$), and White (94.52%, $n = 69$). Five students (6.76%) reported their gender as male. Fourteen students were sophomores (18.92%), and one student was a junior (1.35%). One student reported Asian ethnicity (1.37%), one student identified as Hispanic (1.37%), one student was mixed race (1.37%), and one student listed their ethnicity as "other" (1.37%). Students were mostly pursuing education-related majors. Four students did not report demographic information.

Procedures

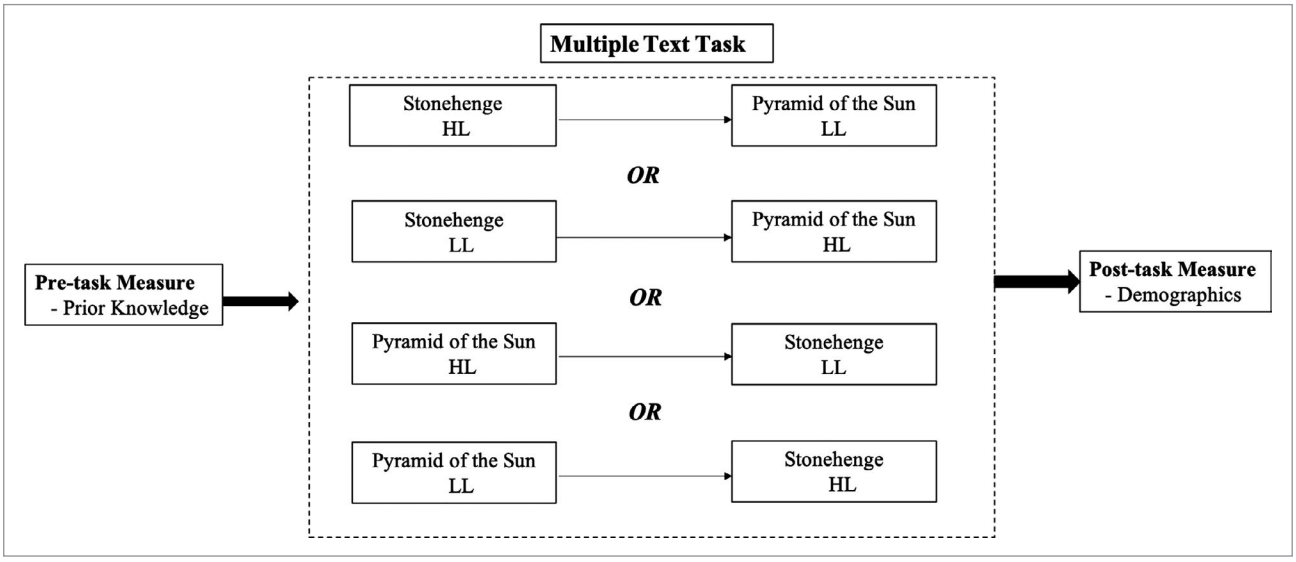
This study had three main parts. First, students completed a variety of individual differences measures, including reporting their prior knowledge related to the topics of this study (i.e., Stonehenge, the Pyramid of the Sun).

Second, students completed a series of two parallel multiple-text tasks. We used a fully crossed, within-subject design, such that students were asked to complete one of two tasks (i.e., a low-level task of responding to multiple-choice items, a high-level task of composing an argument) in association with each of the two topics. For instance, students could have been instructed to read the Stonehenge texts to answer multiple-choice questions and read the Pyramid of the Sun texts to write an argument essay. Both tasks (i.e., low- and high-level) and text topics were presented in counterbalanced order (see Figure 1 for a representation of the procedures). Students were randomly assigned to one of the four conditions, differing in whether they were completing low-level or high-level tasks in association with the Stonehenge or the Pyramid of the Sun topic, in counterbalanced order.

Students read three texts in association with each of the two topics (i.e., Stonehenge, the Pyramid of the Sun). During reading, students were asked to complete a typed strategy elicitation protocol. Such protocols have been found to be an effective means of capturing students' online processing (Muñoz, Magliano, Sheridan, & McNamara, 2006). In the present study, after each paragraph of a text, students were instructed to type whatever they were thinking or doing at the time. Texts were created to be six paragraphs long, with one introductory paragraph presenting a central claim, followed by five paragraphs, each including a single piece of supporting evidence. As such, students typed responses specific to each piece of evidence included in each argument text. Because we were particularly interested in students' processing of text-based evidence, we analyzed only those reports that related to the evidence presented in each text, with students' responses to the introductory paragraph excluded.

Finally, following reading and completion of the typed strategy elicitation protocol, students completed a measure

FIGURE 1
Procedures for the Multiple-Text Task



Note. HL = high-level task (i.e., argument writing); LL = low-level task (i.e., multiple-choice questions).

capturing their strategy use during reading. Then, students completed a performance task (i.e., answered multiple-choice questions or composed an argument), corresponding to their assigned task condition.

The original purpose of this study was to examine differences in students' strategy use while reading to complete two different types of tasks (i.e., answering multiple-choice questions, writing an argument). Nevertheless, as evidence use is often considered to be a key component of effective academic writing, we considered examining students' evidence-related processing strategies and evidence use in writing to be the more theoretically meaningful target for investigation. Therefore, only students' typed strategy reports and performance on the high-level argument task are reported and analyzed. As such, whereas approximately half of the students were randomly assigned to complete the argument task in association with the Stonehenge topic, the remaining students completed the argument task in association with the Pyramid of the Sun topic.

Measures

Prior Knowledge

We assessed students' prior knowledge via a term identification measure. Specifically, students defined six terms related to each of the two topics, or 12 terms in total. Example terms included *Stonehenge*, *trilithon*, *Pyramid of the Sun*, and *radiocarbon dating*. Students' responses were dichotomously scored as correct or incorrect. Total scores ranged from 0 to 6 for each topic. The two authors coded all responses independently. Cohen's kappa inter-rater reliability was .81.

Multiple-Text Task

Students read a set of texts, either about Stonehenge or the Pyramid of the Sun, and wrote an argument about their assigned topic. Specifically, students were instructed to do the following:

> Now, you are going to read three separate texts about Stonehenge/the Pyramid of the Sun. While you read, we're going to stop you at different points throughout each text. We're then going to ask you to type whatever you are thinking or doing at the time. After reading all three texts, you will be asked several questions about your reading process. Then, you will be asked to form an argument about why Stonehenge/the Pyramid of the Sun was built based on the information in the texts.

Texts

The texts used in this study addressed each of the two topics (i.e., Stonehenge, the Pyramid of the Sun). Three texts were provided for each topic. The first text included a brief introduction to each target topic (i.e., "All About Stonehenge"), and the other two texts provided different arguments for why Stonehenge or the Pyramid of the Sun was constructed (e.g., Stonehenge could have been constructed as a burial site or as a center for healing).

The introductory text provided a brief history and general information about each topic (e.g., when Stonehenge was built) and concluded with the explicit recognition that there are various theories regarding why each structure was built. The argument texts were constructed to be isomorphic, or parallel in structure, without making explicit reference to one another. Each argument text included a central claim about why Stonehenge or the

Pyramid of the Sun was built, introduced by a credible author (e.g., Mike Parker Pearson, an archaeologist from the University of Sheffield in England). This claim was then followed by five pieces of supporting evidence. Argument texts were systematically manipulated so they presented two pieces of common evidence, interpreted differently by the conflicting texts, and three pieces of evidence that were unique. For instance, both Stonehenge argument texts agreed that human remains were excavated at the site. At the same time, whereas one of the argument texts interpreted this evidence as demonstrating that "burials were occurring at Stonehenge in all of its years of use," the other argument text interpreted this evidence as Stonehenge specifically attracting ill individuals in need of healing (see Table 1 for a representation of the evidence provided across texts).

The length of the two introductory texts, one for each topic, was 243 words for the Stonehenge topic and 255 words for the Pyramid of the Sun topic, with texts having a Flesch–Kincaid grade level of 11.2 and 12.2, respectively. The length of the four argument texts, two for each topic, ranged from 370 to 381 words, with Flesch–Kincaid grade levels ranging from 12 to 13.8.

Typed Strategy Elicitation

A typed strategy elicitation protocol was adopted, while students were reading to complete the argument task. Before reading about either of the two topics, students read a practice text and reported the strategies they used during reading to familiarize themselves with the protocol. Although asking students to type their strategy use at intermittent points throughout reading has been found to produce fewer reports than traditional think-aloud methods (i.e., that ask students to report their thoughts and behaviors from working memory; Muñoz et al., 2006; Pressley & Afflerbach, 1995), the typed strategy elicitation protocol has been found to be an improvement on self-report methods for measuring strategy use (e.g., strategy checklists), completed following reading. We used the typed strategy elicitation protocol in this study to capture online processing in a somewhat reified or reflective (i.e., cued; van Gog, Paas, van Merriënboer, & Witte, 2005) fashion, in an effort to better understand students' reasoning about evidence in text.

In total, students received 16 strategy elicitation prompts in association with each topic. These were four prompts related to the introductory text and six prompts associated with each of the two argument texts. In the present study, we analyzed only 10 typed reports, related to the 10 pieces of evidence presented in the two argument texts, for each topic. Students' typed reports were coded in two ways, according to their referents (i.e., the object students reported attending to) and the evidence-related processing strategies that they reflected.

We identified four primary referents in students' typed reports: students referring to (1) a single text (i.e., the segment they were reading, segments within the same text), (2) multiple texts (i.e., segments pertaining to texts other than the one they were reading), (3) their prior knowledge, or (4) themselves as learners. For instance, we coded a report like "more evidence as to why there were sacrifices" as referring to a single text, whereas we coded a report like "This is very similar evidence that was used to support the burial site theory because there were bodies there," as referring to multiple texts. We assigned multiple codes when a report referred to more than one object. For instance, we coded reports like "It is interesting how Stonehenge was actually for the dead not living. I always thought it was for some ritual to cure people," as reflecting two referents. Specifically, we coded the first part of this report as referring to a single text because it drew on content from the text that this student was reading at the moment. We coded the second part of this report as referring to prior knowledge because it referenced this student's prior understanding of why Stonehenge was built. See Table 2 for detailed explanations and examples for each referent.

We coded students' evidence-related processing strategies in two ways: according to the level of strategy use that each report corresponded to (i.e., surface, deep, metacognitive, affective) and according to the specific strategy that was reflected. We distinguished surface- and

TABLE 1
Representation of Evidence Provided Across Texts

	Stonehenge		Pyramid of the Sun	
Claim	**Burial site**	**Center for healing**	**Temple**	**Burial site**
Evidence 1	*Human remains*	Stone chips	Alignment	*Human remains*
Evidence 2	Religious object	*Human remains*	*Chamber myth*	*Chamber myth*
Evidence 3	*Village nearby*	Pilgrimage	Charcoal	Tombs
Evidence 4	River Avon	*Village nearby*	*Human remains*	Avenue
Evidence 5	Structure	Historical account	Red paint	Structure

Note. Italics denote pieces of evidence that were commonly presented across texts.

TABLE 2
Coding Scheme for Referents

Category	Description	Example
Single text	Students referring to the text they were reading at that moment	"Radio carbon does not interest me."
Multiple text	Students referring to information from more than one text that they read	"If this was thought to be a place of healing, why are so many people buried there? That is why this theory makes more sense."
Prior knowledge	Students referring to their previous knowledge about the topic	"I did not know the ancient English people buried stuff with dead people for the afterlife. That reminds me of Egypt."
Themselves	Students referring to themselves as learners	"Not really thinking much of it"

deep-level strategies using Dinsmore and Alexander's (2012) conceptualization. In particular, surface-level strategies were cognitive activities directed only at the basic level of information encoding and included strategies such as restating information from texts. Deep-level strategies reflected cognitive efforts at information transformation in some way, such as forming cross-textual connections or evaluating evidence in texts. Affective strategies expressed emotions or other personal reactions to the information processed. Metacognitive strategies included students' reflections on the cognitive strategies they engaged, reports of their comprehension successes or failures, and considerations of themselves as learners.

We identified six categories of specific strategies used: restating information in texts, with limited transformation (e.g., repeating information from the text); generating elaborative questions (i.e., asking interpretive questions about text content); elaborating text-based content (i.e., explaining text-based information using prior knowledge); evaluating evidence (i.e., judging text-based information using logic or personal standards of coherence); reporting a personal reaction (i.e., expressing emotions or opinions related to text-based information); and demonstrating self-regulation (i.e., monitoring and controlling reading processes). See Table 3 for detailed descriptions and examples of each category.

Self-Reported Processing Strategies

After students completed reading the three texts associated with each topic, they completed a self-report measure assessing their strategy use during reading. Specifically, students rated their use of 16 strategies on a 5-point Likert-type scale, ranging from 1 (*strongly disagree*) to 5 (*strongly agree*). Items were drawn from Anmarkrud and Bråten

TABLE 3
Coding Scheme for Processing Strategies

Category	Description	Example
Restating	Repeating or paraphrasing information directly from text(s)	"Merlin, a wizard said that Stonehenge was an area of healing."
Elaborative questions	Asking interpretive questions about information in text(s)	"I wonder if they are ever going to be able to study underneath the other pyramids. Is it ethical to be finding and looking for the remains of ancient people?"
Elaboration	Using prior knowledge to explain information provided in text(s)	"This symbolism is very sweet and definitely fits in with the cultural feel of Mesoamerica, although, I think that the fact that the men and women were separated speaks a lot about how the culture from Mesoamerica to Latin America/South America has changed in terms of unity in the family. However, they were royal so that explains the extravagance."
Evaluation	Assessing text-based information using certain standards, such as logic	"Using DNA analysis is very persuasive on the perspective…."
Personal reaction	Stating personal opinions or reactions toward information from text(s)	"This is literally a pyramid person. It would honestly be super cool to go into it."
Self-regulation	Regulating reading processes by self-monitoring and controlling comprehension	"I am paying attention more to this passage because I feel like there is going to be a quiz, just like the last one. I now know what to expect from this."

(2009) and tapped two levels of processing: surface and deep. Whereas surface-level strategies reflected students' use of memorization, deep-level strategies demonstrated students' efforts to organize and elaborate on the information they read and to monitor their reading process. We modified items to reflect multiple-text comprehension. Although defined somewhat distinctly, the deep-level strategies coded for in students' typed strategy reports and the deeper strategies captured via Anmarkrud and Bråten's self-report measure both emphasized students' analytic and evaluative engagement with and manipulation of information from texts. These were contrasted to students' surface-level processing or restating of information from texts. Sample items for the surface strategy subscale included "I tried to memorize as much as possible," and "I made sure that I remembered the most important things." Sample items from the deep-level strategy subscale included "I tried to make all the content in the texts fit together," and "I tried to find connections between the different parts of the texts." Students' self-reported strategy use scores, on average, were 3.31 (SD = 0.78) for surface strategy use and 3.66 (SD = 0.62) for deeper strategy use. The Cronbach's alpha reliabilities for surface and deeper strategy use subscales were .76 and .72, respectively. Correlations between students' scores on both strategy use subscales and other processing and performance variables examined in this study are presented in the Appendix.

Argument Coding

We also coded the arguments composed by students in association with each topic in two ways. First, we totaled the number of accurate pieces of evidence used in students' written responses. In other words, students received 1 point when they included one piece of evidence from one particular text to support its correspondent claim. Students received 0 points when they used evidence inaccurately (e.g., using evidence from one argument text to support a claim from the opposing argument text, with no justification provided). The total number of pieces of evidence that students could include in their written responses ranged from zero to eight.

In addition, given that the provision of opposing reasons and their rebuttal has been considered to be indicative of high-quality argumentative writing (e.g., Anmarkrud et al., 2014; De La Paz & Felton, 2010; Reznitskaya et al., 2009), we further coded students' specific use of the two pieces of common evidence, presented across the two argument texts. We examined such evidence use because we were interested in how students analyzed common evidence (i.e., evidence commonly introduced but differentially interpreted across texts) in their writing. We considered the ability to attend to and reconcile such conflicting interpretations of evidence to be essential for students' reasoning about controversial topics described across multiple texts (Kobayashi, 2014; Mason, Ariasi, & Boldrin, 2011; Thomm & Bromme, 2016). Moreover, because common evidence was, by definition, evidence that was discussed across both of the argument texts that students read, we expected them to specifically pay attention to this evidence and its interpretation. As such, we examined both quantity-related (i.e., total evidence use) and quality-related (i.e., common evidence use scores) indicators of evidence use in writing. In other words, we examined both the volume of evidence that students included in their writing and how effective students were in writing about the common evidence introduced across texts.

We used a 4-point scale to code students' use of common evidence in writing. Specifically, students received a 0 if they included none of the common evidence in the responses that they composed. A 1 was given when one piece of common evidence was used to support just one explanation for the construction of Stonehenge/the Pyramid of the Sun. This reflected students essentially ignoring the fact that this same evidence was variably interpreted in a different text to support a conflicting explanation for each structure's construction. A 2 was given when common evidence was provided to support each explanation for why Stonehenge/the Pyramid of the Sun was built, with no further justification provided. Again, this score reflected students essentially ignoring that common evidence was distinctly, and potentially incompatibly, interpreted across conflicting texts. Finally, a 3 was assigned when students presented common evidence to support both explanations for the construction of Stonehenge/the Pyramid of the Sun, while at the same time offering some justification or reconciliation regarding how this evidence was used. Total scores for common evidence use ranged from 0 to 3.

Results

Research Question 1: Students' Evidence-Related Processing Strategies

For our first research question, we examined the evidence-related processing strategies that students reported using while reading. A total of 754 reports were generated, for which 854 referents were identified (i.e., on occasion, a strategy was directed toward more than one referent). Overall, students mostly referred to a single text in the typed responses that they produced (n = 740, 86.65%), followed by referring to prior knowledge (n = 52, 6.09%), multiple texts (n = 46, 5.39%), and to themselves as learners (n = 14, 1.64%). The majority of strategies that students reported using reflected deep-level processing (n = 446, 52.72%), followed by surface-level (n = 286, 33.81%), affective (n = 98, 11.58%), and metacognitive (n = 16, 1.89%) strategies reported.

In terms of the specific evidence-related processing strategies reported, on average, students primarily reported

engaging in restating (M = 3.13, SD = 3.63), followed by reports of evidence evaluation (M = 1.81, SD = 2.04) and elaboration (M = 1.72, SD = 1.95) and the asking of elaborative questions (M = 1.51, SD = 1.99). Students less commonly reported personal reactions to the information read (M = 0.53, SD = 0.98) and rarely engaged in self-regulation (M = 0.40, SD = 0.69). Although restating was, individually, the strategy that students reported using most often, when strategies were grouped as surface or deep level in nature, students were found to engage more deep-level than surface-level strategies overall. In part, this may be due to the variety of deep-level strategies that students reported engaging (e.g., elaboration, evaluation), whereas restating was the primary surface-level strategy that students reported using.

Students' engagement of strategies related to evidence evaluation was of particular interest to us in this study. Indeed, 61.54% of students (n = 48) reported using at least one strategy reflective of evidence evaluation during processing. At the same time, some of the evaluations that students reported seemed to be fairly superficial or surface level in nature. For instance, one student reported, "I don't really buy into this point," after reading one piece of evidence presented, whereas another student's evidence evaluation was "I think that is great evidence for this thoery [theory]." As demonstrated in these responses, such instances of evaluation may reflect somewhat superficial judgments of evidence quality rather than reflective and epistemically based considerations.

Research Question 2: Students' Evidence Use in Writing

For our second research question, we examined students' evidence use in writing, including the total amount of evidence provided in support of claims and the quality of students' common evidence use. Across the two topics, students used an average of 1.96 (SD = 1.39) pieces of evidence to support their claims, reflecting a relatively small volume of evidence incorporated. Students' average scores for common evidence use were 0.72 (SD = 0.93), indicating that they commonly failed to use such evidence, even when this evidence was repeatedly introduced across texts. Moreover, even in cases when common evidence was used, this was most often employed to support only one explanation for Stonehenge's or the Pyramid of the Sun's construction, with no seeming recognition that this evidence could also be variably interpreted to support a conflicting conclusion.

Research Question 3: Association Between Typed Evidence Processing Strategies and Use of Evidence in Writing

For our third research question, we examined the association between students' reported evidence-related strategy use during reading and their evidence use in writing. Correlations among key variables are provided in Table 4. We used multiple regression analysis to predict the total amount of evidence used in students' writing and the quality of students' common evidence use, based on strategies reported through the typed strategy elicitation protocol.

Total Evidence Use

We ran a multiple regression to predict the total amount of evidence that students accurately used in composing their written responses based on multiple texts. Prior knowledge and topic were controlled for in step 1, with evidence-

TABLE 4
Correlations Between Evidence-Related Processing Strategies and Evidence Use in Writing

Strategies and performance	1	2	3	4	5	6	7	8
Evidence-related processing strategies								
1. Restating	—							
2. Elaborative question	−.40***	—						
3. Evaluation	−.53***	−.06	—					
4. Elaboration	−.32**	−.12	−.08	—				
5. Personal reaction	−.39***	.17	.21	−.07	—			
6. Self-regulation	−.27*	.19	.12	.10	.09	—		
Evidence use in responses								
7. Total evidence use	.00	−.15	.28*	.04	.01	.00	—	
8. Common evidence use	−.10	−.08	.30**	.13	.17	.05	.64***	—

*The correlation is significant at the .05 level (two-tailed). **The correlation is significant at the .01 level (two-tailed). ***The correlation is significant at the .001 level (two-tailed).

related processing strategies entered at step 2. However, the overall model predicting the volume of evidence included in students' written responses was not statistically significant ($p = .09$). A post hoc power analysis using G*Power, with $\alpha = .05$, eight predictors, a sample size of 78, and a determined effect size of $f^2 = 0.22$, reflecting a medium-large effect, indicated that we had sufficient power to identify any statistically significant variance in the volume of evidence use explained by the model ($1 - \beta = .81$).

Common Evidence Use

We ran a second multiple regression model to predict the quality of students' common evidence use. In this model, we were specifically interested in examining students' writing about evidence that was commonly cited but differently interpreted across texts. Prior knowledge and topic were again controlled for in step 1, with evidence-related processing strategies entered in step 2. The overall model was statistically significant, $F(8, 66) = 4.13$, $p < .01$, $R^2_{adj} = .25$, reflecting a large effect. Text topic ($\beta = -0.26$, $p < .05$), students' use of restating ($\beta = 0.81$, $p < .01$), evaluating ($\beta = 0.78$, $p < .001$), and elaborating ($\beta = 0.56$, $p < .01$) strategies, as well as reports of personal reactions to text-based evidence ($\beta = 0.38$, $p < .01$), were statistically significant predictors in the model (see Table 5 for a model summary).

Given that text topic was a significant predictor in the model, we further ran separate regression analyses for each of the two topics. Descriptive statistics, by topic, are presented in Table 6. When we analyzed topics separately, only the model predicting students' common evidence use when writing about the Pyramid of the Sun topic was statistically significant, $F(7, 34) = 3.05$, $p < .05$, $R^2_{adj} = .26$, reflecting a large effect. Students' use of restating ($\beta = 0.76$, $p < .05$), evaluating ($\beta = 0.91$, $p < .01$), and elaborating ($\beta = 0.59$, $p < .05$) strategies were statistically significant predictors in the model (see Table 7 for a model summary). The regression model predicting common evidence use for the Stonehenge topic was not statistically significant ($p = .19$).

Discussion

Students' Evidence-Related Processing Strategies

For our first research question, we examined the referents and evidence-related processing strategies reflected in students' typed strategy reports. Results showed that when reasoning about evidence, students mostly referred to the evidence included within the specific text that they were reading (86.65%) and rarely corroborated evidence across multiple texts (5.39%). Moreover, students infrequently referred to their prior knowledge during processing (6.09%), although this may be attributable to students' limited degree of prior knowledge, overall. These results both parallel and diverge from findings in prior work. This study echoes prior work finding students to have limited performance on tasks requiring the integration of multiple texts (e.g., Gil et al., 2010; List et al., 2019) and is specifically aligned with prior work that has found students to refer to multiple texts infrequently when thinking

TABLE 5
Model Predicting Common Evidence Use Across Topics

					Correlations		
Predictor	***B***	**Standard error**	**β**	**Significance**	**Zero-order**	**Partial**	**Part**
Step 1: Control variables							
Prior knowledge	0.02	0.07	0.03	.80	.06	.03	.03
Topic	−0.68	0.29	−0.26	.02	−.24	−.28	−.24
Step 2: Evidence-related processing strategies							
Restating	0.29	0.09	0.81	.00	−.10	.37	.32
Elaborative question	0.20	0.11	0.31	.06	−.08	.23	.19
Evaluation	0.50	0.12	0.78	.00	.30	.46	.43
Elaboration	0.37	0.11	0.56	.00	.13	.38	.34
Personal reaction	0.50	0.16	0.38	.00	.17	.36	.31
Self-regulation	0.11	0.20	0.06	.58	.05	.07	.06
Overall: $F(8, 66) = 4.13$, $p < .01$, $R^2_{adj} = .25$							

Note. Effect sizes for zero-order correlations: ±.10 = small effect size; ±.30 = moderate effect size; ±.50 = large effect size.

TABLE 6
Descriptives for Variables of Interest

Strategies and performance	Stonehenge (*n* = 35)		Pyramid of the Sun (*n* = 43)		Total (*N* = 78)	
	Range	*M* (*SD*)	Range	*M* (*SD*)	Range	*M* (*SD*)
Prior knowledge	7.00	1.46 (1.72)	7.00	2.70 (2.22)	7.00	2.14 (2.09)
Evidence-related processing strategies						
Restating	10.00	3.83 (3.94)	10.00	2.56 (3.28)	10.00	3.13 (3.63)
Elaborative question	7.00	1.49 (2.08)	7.00	1.53 (1.93)	7.00	1.51 (1.99)
Evaluation	8.00	1.51 (1.85)	7.00	2.05 (2.17)	8.00	1.81 (2.04)
Elaboration	9.00	1.54 (2.08)	8.00	1.86 (1.86)	9.00	1.72 (1.95)
Personal reaction	5.00	0.37 (1.00)	3.00	0.65 (0.95)	5.00	0.53 (0.98)
Self-regulation	3.00	0.40 (0.78)	2.00	0.40 (0.62)	3.00	0.40 (0.69)
Multiple-text processing strategies						
Surface-level reading strategies	2.57	3.34 (0.72)	4.00	3.29 (0.83)	4.00	3.31 (0.78)
Deep-level reading strategies	2.89	3.54 (0.60)	3.56	3.76 (0.62)	3.56	3.66 (0.62)
Performance						
Total evidence use	5.00	2.03 (1.38)	5.00	1.90 (1.41)	5.00	1.96 (1.39)
Common evidence use	4.00	1.79 (1.47)	4.00	1.17 (1.10)	4.00	1.44 (1.31)

Note. *M* = mean; *SD* = standard deviation.

TABLE 7
Model Predicting Common Evidence Use for the Pyramid of the Sun Topic

Predictor	*B*	Standard error	β	Significance	Correlations		
					Zero-order	Partial	Part
Step 1: Control variables							
Prior knowledge	−0.03	0.08	−0.05	.74	.14	−.06	−.04
Step 2: Evidence-related processing strategies							
Restating	0.26	0.11	0.76	.02	−.14	.38	.32
Elaborative question	0.19	0.14	0.34	.16	−.20	.24	.19
Evaluation	0.46	0.14	0.91	.00	.43	.50	.46
Elaboration	0.34	0.14	0.59	.02	.15	.40	.34
Personal reaction	0.26	0.20	0.23	.21	.12	.22	.17
Self-regulation	0.38	0.24	0.21	.13	.29	.26	.21
Overall: $F(7, 34) = 3.05$, $p < .05$, $R^2_{adj} = .26$							

Note. Effect sizes for zero-order correlations: ±.10 = small effect size; ±.30 = moderate effect size; ±.50 = large effect size.

aloud during task completion (Wolfe & Goldman, 2005). As reflected in this study, students rarely corroborated information across texts, even when the texts presented overlapping and directly conflicting interpretations of the same information.

Differing from results found in some prior work, students in this study were limited in the extent to which they drew on prior knowledge during processing. In a study of middle school students, reading about the decline of the Roman Empire, Wolfe and Goldman (2005) found that

almost half of students' think-aloud utterances were based on their prior knowledge. This difference in the prevalence of prior knowledge engagement during reading between our results and findings from Wolfe and Goldman may be explained in a variety of ways. First, Wolfe and Goldman found 70% of prior knowledge utterances to reflect irrelevant associations, not contributing to students' understanding of texts' content. Such irrelevant associations may be less commonly reported by students as they develop. That is, this decrease in the citation of irrelevant associations may be attributable to the sample examined in Wolfe and Goldman's study constituting middle school students, whereas the students in our sample were undergraduates. Second, contrary to the think-aloud protocol in Wolfe and Goldman's study, students' in the present study were asked to type their thoughts or behaviors instead of orally reporting them. Prior research (Muñoz et al., 2006) has found that students tend to report more strategy use when asked to speak their thoughts or behaviors as compared with typing these. As such, the adoption of a typed strategy elicitation protocol in the present study may have restricted students' reports of prior knowledge use, in addition to limiting reports of strategy use, more generally. Third, students in Wolfe and Goldman's study had received six to eight weeks of instructions on the topic of the Roman Empire prior to participating in the study. This may have facilitated their use of prior knowledge while reading. Nevertheless, in the present study, students did not receive any instruction on the texts' topics and performed poorly on the prior knowledge measure. These limitations in prior knowledge may also have constrained students' reports of strategy use during processing.

Moreover, different strategy types, as documented in prior work, were also identified in this study. In particular, students reported using surface- and deep-level strategies alongside each other during reading. According to Dinsmore and Alexander (2012), surface-level strategies are focused on the basic encoding of information, whereas deep-level strategies require the transformation of information. In the present study, 52.72% of the strategies identified were deep level in nature, whereas 33.81% of strategies were classified as surface level, suggesting that these were jointly used by students to support comprehension. At the same time, metacognitive strategy use, including comprehension monitoring, was found to be reported only to a limited extent; this finding is also consistent with prior work (Stadtler & Bromme, 2007).

Looking at the specific strategies that students reported engaging during reading, these most commonly involved restating evidence (M = 3.13, SD = 3.63), with evidence elaboration (M = 1.72, SD = 1.95) and evaluation (M = 1.81, SD = 2.04) invoked to a somewhat more limited extent. These prevalent reports of restating while processing text-based evidence can be attributed to students' limited prior knowledge about the topics of the task. Such limits in prior knowledge may have restricted the degree of elaborative strategies that students may have been able to engage during reading. Indeed, limits in prior knowledge have been found to result in more superficial or surface-level processing, particularly in the initial stages of learning (Dinsmore, Hattan, & List, 2017; Novak, 2010).

A substantial number of students in this study engaged in evaluation, relative to the proportion of students who have been found to do so in prior work (Chinn & Brewer, 2001; List et al., 2017). Still, the types of evaluations that many students reported may best be described as plausibility judgments (e.g., "I believe this because it seems to make sense"), rather than deliberate and systematic efforts at evidence evaluation. Lombardi, Sinatra, and Nussbaum (2013) defined plausibility judgments as implicit and automatic judgments of the potential truthfulness of information and suggested that these commonly occur when students reason about conflicting information. A direction for future work is to examine differences in students' more superficial (i.e., plausibility based) vis-à-vis deep-level (i.e., explicit, effortful) evidence evaluation and the extent to which these are differentially associated with task performance.

The superficial nature of students' reports of evaluation-related strategy use can also be attributed to the typed strategy elicitation protocol used in the present study. Specifically, when cueing students to report their processing strategies, we did not explicitly ask them to specify the particular evaluation criteria that they were relying on in making their determinations of evidence quality. This may have resulted in students simply reporting that they found evidence to be of high quality or not, without typing out any further justifications. Future studies should consider using prompts or instructions that explicitly ask students to report their rationale when judging evidence quality to better assess the nature of students' evaluation strategy use. Indeed, this has been done to some extent in prior work. For instance, Gerjets, Kammerer, and Werner (2011) found students' text evaluations more generally, rather than evidence evaluations specifically, to increase when students received explicit instructions to evaluate websites during information search. As another note, the superficial nature of students' evidence evaluations suggests the challenges of dichotomizing strategies as surface or deep level in nature. Although surface- and deep-level strategies have commonly been used to distinguish students' processing (Cromley, Snyder-Hogan, & Luciw-Dubas, 2010; Dinsmore & Alexander, 2012), more recent investigations of strategy use during students' learning from multiple texts have suggested that other strategy categorizations may prove more fruitful (e.g., intertextual, metacognitive; Afflerbach & Cho, 2009; Anmarkrud et al., 2014; List, 2020). Future work should examine the association between students' more specific strategy use and task performance.

Moreover, certain aspects of the way evidence was introduced to students in this study may have helped or hindered their evaluation process. First, evidence provided to students was introduced through fairly short and simple texts. This may have eased the evaluation process for students by reducing comprehension demands; at the same time, students were provided with fairly limited information that they could use in making determinations of evidence quality. Second, the evidence that students were provided was associated only to a limited extent with their prior knowledge, likely making the evidence evaluation process a more challenging task for learners. Third, evidence was included in the texts used in this study based on its ability to be comparably and succinctly introduced across texts and to be uniquely or commonly discussed across texts, rather than based on its quality, per se. This may have created a particularly challenging evaluation context for learners. Fourth, in this study, students were presented with directly conflicting interpretations of the same evidence across argument texts, attributed to comparably expert authors. Given Stadtler and Bromme's (2014) work suggesting that students compare author credibility to resolve conflicts, when unable to do so directly based on their prior knowledge, the comparability of author expertise likely contributed to making evidence evaluation a particularly challenging task for learners in this study. All the more so, because rather than being able to cognitively represent the conflicting evidence distinctly, in essence ignoring any conflict, students in this study were forced to grapple with the fact that the evidence presented overlapped but was differentially interpreted across conflicting texts.

Most of all, it is likely the case that students were not aware of the relevant criteria to use in evaluating evidence and how these ought be applied. Indeed, to be able to evaluate evidence successfully, students need to possess disciplinary understanding and knowledge of methodology and research design (Duncan, Chinn, & Barzilai, 2018), competencies far outside the scope of expertise for our sample of domain novices. Although presenting novice learners with common evidence, incompatibly interpreted across expert texts, was our intention, this study highlights the difficulties that students experience when asked to reason about comparably trustworthy texts presenting conflicting information.

Students' Evidence Use in Writing

For our second research question, we examined students' use of text-based evidence in writing. Two indicators of evidence use were considered. The first was a quantitative measure reflecting the total amount of evidence that students used in the written responses that they composed. The second was a rating scale that aimed to capture the quality of students' evidence use, with quality in this study reflecting the appropriate use of evidence (i.e., in support of correspondent claims) and its corroboration (i.e., recognition of its disparate interpretation) across texts.

In terms of the total volume of evidence use, students included an average of 1.96 (*SD* = 1.39) pieces of evidence in their written responses, out of a possible eight. This indicates that students included only a minority of the evidence provided across texts to support their chosen position regarding the construction of Stonehenge or the Pyramid of the Sun. This is concerning given that the provision of evidence to support or justify claims has been identified as foundational to argument writing (De La Paz et al., 2012; Du & List, 2020; Iordanou & Constantinou, 2014). This paucity in the volume of evidence that students included in their written responses may be explained in three ways.

First, students may not have known the appropriate or sufficient amount of evidence to include in their writing (i.e., signaling a lack of discourse knowledge; Olinghouse & Graham, 2009; Scardamalia & Paris, 1985). Also, students may have lacked more general knowledge regarding what writing about evidence entailed, including the need to select and organize evidence from texts and to describe and elaborate evidence through writing—all processes making evidence-based writing an overall demanding task. Second, students may have been less adept at using evidence in writing because of limitations in their prior knowledge. That is, students' limited knowledge of the topics examined in this study may have made evidence recall more difficult for students during writing or may have made writing about evidence a greater challenge for learners. Finally, students' limited evidence use in writing could be attributed to motivational factors, such as students' limited interest or the limited value that students assigned to the writing task. Indeed, as noted by Bråten, Brante, and Strømsø (2018), one of the reasons why students do not sufficiently engage with multiple-text tasks could be that they do not find such tasks to be relevant to their own areas of study. As a manipulation check, students in this study were asked to respond to two questions—How interested were you in this study? How much did you pay attention during this study?—on a Likert-type scale of 0–5 at the end of the study. Students' average ratings of these two questions were 2.72 (*SD* = 1.17) and 3.46 (*SD* = 1.17), respectively. This suggests that students had at least a moderate degree of engagement with this study. Nevertheless, it may be the case that students were reluctant to expend much effort on writing task completion. Students may be more motivated for task completion when they are asked to write as part of a class or for a graded assignment. In such instances, students may be expected to include more evidence in their writing.

We assessed the quality of students' evidence use by examining how they wrote about common evidence, presented in parallel across argument texts but differentially

interpreted across these. Findings examining common evidence use determined that students most often only used a single piece of the common evidence to support a single claim. This means that students ignored evidence, even when this was commonly cited by conflicting texts, and that students used such conflicting evidence, with no explicit acknowledgment of the associated conflict. This is a concerning finding given the frequency with which students are confronted with conflicting information presented across multiple texts (Kienhues, Stadtler, & Bromme, 2011; Mason et al., 2011).

At the same time, ignoring conflicting evidence when it is presented across texts has been documented as a commonly occurring phenomenon. According to the content–source integration model (Stadtler & Bromme, 2014), which describes the strategies that students undertake when presented with conflicting information, ignoring conflict is a frequent strategy used to achieve cognitive coherence. As such, in this study, students either may have deliberately ignored the conflicting nature of the evidence introduced or may have been unwilling or unable to resolve conflicts arising across texts, even when recognizing such conflicts as present.

Finally, an interesting pattern that emerged was that more students (10.67%) received a score of 3 according to the common evidence use rubric than students receiving a score of 2 (1.33%). This pattern of scores reflects the fact that when students associated a single piece of common evidence with the two conflicting arguments presented across texts, they more often elaborated or reconciled this conflict than simply leaving it unmentioned. This pattern seems to reflect the fact that when presented with conflicting evidence across texts, learners choose either to ignore one aspect of this conflict or to engage with this conflict in an effort to achieve resolution. It was relatively unusual for students to only recognize a conflict and then to do nothing to resolve it. At the same time, being able to cognitively represent conflicting information, sans further resolution, may be an important competency for students to develop. This competency may be especially necessary when students learn about controversial topics, about which there is substantial expert disagreement and no valid way for novice learners to achieve resolution. Given the number of unsettled issues among experts on various topics across domains, teaching students how to reason about such unsettled issues and represent these through writing constitutes a key direction for future work.

Association Between Evidence-Related Strategies and Use of Evidence in Writing

For our third research question, we examined the association between evidence-related processing strategies reported by students during reading and their ultimate evidence use during writing. Although none of the processing strategies reported were significantly associated with the total amount of evidence included in students' written responses, we found that strategies, such as restating, evaluating, and elaborating, were statistically significant predictors of the quality of students' evidence use in writing. Indeed, evaluation and elaboration have been found to be deep-level strategies associated with multiple-text task performance in prior work (Bråten et al., 2009; De La Paz et al., 2012).

At the same time, in this study, we also found students' more surface-level processing to be significantly and positively associated with multiple-text writing performance. This stands in contrast to prior work (Bråten & Strømsø, 2011) that found surface-level processing to be negatively associated with performance. In part, this difference in outcomes may be attributable to differences in how surface-level processing has been operationalized across studies. In Bråten and Strømsø's (2011) work, surface-level processing was examined as students' engagement in information accumulation (i.e., gathering as much information as possible from texts); as a contrast, in our study, surface-level processing was captured via students' paraphrasing of text-based evidence during reading. Although a comparatively superficial strategy, vis-a-vis evaluation, paraphrasing in this study may nevertheless have conferred a number of benefits for learners. In this study, as in prior work, paraphrasing may have helped students restate text-based information using more accessible language or may have cued students to engage other strategies related to comprehension (McCarthy, Guess, & McNamara, 2009). Indeed, paraphrasing has been identified as a particularly useful strategy for low-knowledge learners (McNamara, 2004). More generally, the use of paraphrasing in this study may have drawn students' attention to the text-based evidence provided and may have served as a precursor to more sophisticated or deeper level processing. That is, any evidence-related strategy reporting in this study, even if this involved only the paraphrasing of text-based evidence, was likely preferable to no such strategy reporting at all. Nevertheless, although the texts used were manipulated to be similar in length and reading level and to present evidence in a parallel fashion across the two topics, we also found text topic to be a significant predictor for students' common evidence use in writing. Therefore, caution should be taken when generalizing this finding.

When we considered performance on each topic separately, only the model predicting common evidence use for the Pyramid of the Sun topic was statistically significant. In part, this may be due to the reduction in sample size that resulted from our running of separate regressions for each of the two topics. Indeed, to obtain an effect size of $f^2 = .25$, corresponding to a medium effect, with a power of .80, with seven predictors involved, a sample size of 65 participants would have been needed. However,

the sample size when examining each topic separately was only 35 students. Future studies should consider including more participants to allow for more robust cross-topic comparisons.

Another explanation for the statistical significance of the model predicting students' performance on the Pyramid of the Sun topic vis-à-vis the Stonehenge topic is that the two arguments introduced for the Stonehenge topic seemed to be more balanced than those introduced in association with the Pyramid of the Sun topic. That is, when asked to take a stance on each of the focal controversies, 78.57% of students considered the Pyramid of the Sun to be constructed as a burial site, whereas 30.95% of students believed it was constructed as an ancient temple, including students who supported both sides. At the same time, for the Stonehenge topic, 57.58% of students considered Stonehenge to be constructed as a burial site, whereas 75.76% of students considered it to be constructed as a center for healing. This suggests that students considered one of the argumentative Pyramid of the Sun texts to be more convincing than the other but considered the two argumentative Stonehenge texts to be more balanced. This imbalanced perception of texts' convincingness may be attributable to at least two factors.

First, students' stronger endorsement of burial as the explanation for the Pyramid of the Sun's construction may have been driven by their knowledge of the Pyramids of Giza, used for royal tombs. Indeed, some students mentioned the Egyptian pyramids in their typed responses when reading about the Pyramid of the Sun (e.g., "this makes a lot of sense to me because it is in line with what I learned about the Pyramids of Giza in Egypt"), and this possibility was suggested in the burial texts (i.e., "Finally, the structure of the Pyramid of the Sun mirrors other burial practices in the ancient world, most notably the pyramids in Egypt"). Second, because both sets of argumentative texts reported that human remains were found at the site of the Pyramid of the Sun, students may have been particularly inclined to take for granted that individuals were, in fact, buried there and to (incorrectly) infer that the Pyramid of the Sun was used both for religious rituals and for burial. This was reflected in responses such as "I believe that they could hold ceremonies inside of the temple while still leaving it as a place for the dead to be buried in peace considering it was also used for religious reasons and sacrifice." Such responses are consistent with Stadtler and Bromme's (2014) findings that a common way for individuals to manage conflicts appearing across texts was to make (often inaccurate) inferences about these to reconcile or resolve them. This seemed to be an approach adopted by some students in this study.

Regardless of the explanation, the difference in texts' convincingness across topics, as manifest in students' ultimate topic stances, may have resulted in the differences in performance identified across topics. That is, the differential characteristics of the Pyramid of the Sun texts may have contributed to the strategies used during their processing being more associated with performance (i.e., the deployment of particular strategies may have better allowed students to distinguish among these texts), wherein the use of different types of strategies may have been less differentially effective when applied to the more balanced Stonehenge texts. Nevertheless, this possibility requires further examination. More generally, there is a need to examine more text topics, beyond these, in future analyses to further establish the association between evidence-related processing strategies and evidence use in writing.

Contribution to Prior Work

In this study, we examined students' evidence use when confronted with a common, yet formidable, challenge: reasoning about conflicting information introduced by authors of comparable expertise. We examined such reasoning specifically as it pertains to students' processing of unique and overlapping evidence presented across texts and to students' use of such evidence in their writing. In analyzing students' evidence-related reasoning, this study contributes to the extant literature in at least two primary ways. First, we provide a framework for examining the nature of students' evidence-based reasoning. That is, we suggest that such reasoning includes the processing and evaluation of evidence during reading and the elaboration, integration, and reconciliation of such evidence during writing. Such an analysis stands in contrast to prior work that has focused on students' evidence-based reasoning only to a limited extent and has viewed students' evidence evaluation during reading as a process that is fairly separate from students' evidence use in writing. In this study, we demonstrated the connection between these two processes (i.e., reading and writing).

Second, this study contributes to work examining students' learning from conflicting texts, such as those commonly encountered when students read information online. Although prior studies examining students' learning from conflicting texts have attributed such conflict to sources varying widely in quality or expertise (e.g., sources with benevolent or commercial intent; Kammerer, Meier, & Stahl, 2016), this study is distinct in examining a unique instance of conflict-related reasoning: students' reasoning about conflicting evidence presented by experts of comparable trustworthiness. We found such reasoning to be quite challenging for students and to reflect the use of strategies previously only theoretically specified (Stadtler & Bromme, 2014). Although we found instances of students ignoring expert conflict and attempting to justify or reconcile such conflict, potentially in an inappropriate fashion, we found relatively limited evidence that students are comfortable with understanding and writing about conflicting information that reflects differences in expert

perspectives, rather than a conflict that can easily be understood and resolved. This suggests that a key area for future instruction is helping students understand controversial topics that reflect not a difference between expert and lay opinion (e.g., climate change, vaccinations) but rather topics that reflect a lack of expert consensus (e.g., opioid prescription). That is, the strategies that students ought to be expected to draw on when presented with conflicting information from comparably expert texts remain unclear.

Limitations

Despite the strengths of the present study, several limitations must be acknowledged. First, we used a typed elicitation protocol (i.e., asking students to type their thoughts or behaviors at designated points within a text). Although this methodological approach provides an efficient way of capturing students' online processing and segmenting students' strategy reports, this methodology may have introduced some limitations. To start, this specific mode of think-aloud (i.e., typing) may have affected students' reports of strategy use. In a comparison of students' oral and typed reports during reading, Muñoz et al. (2006) found that students reported using a higher number of strategies overall and the higher engagement of certain strategies in particular (i.e., paraphrasing, bridging inferences), when asked to orally report their thoughts and behaviors rather than typing these. As such, the use of a typed elicitation protocol in the present study may have restricted or otherwise modified students' reports of strategy use. Likewise, asking students to type, rather than spontaneously think aloud, may have resulted in students further reifying their thoughts or behaviors in a more formal fashion (van Gog et al., 2005). This may have resulted in differences in students' descriptions of strategic processing and, therefore, in our abilities to identify and classify these.

Additionally, although in our view, a typed strategy elicitation was preferable to self-report measures of strategy use (e.g., a strategy rating scale), it is important to note that in the present study, students were cued to report their thoughts and behaviors at researcher-designated points throughout reading. This may not have fully captured all of the strategies that students deployed, as can be accessed via continuous think-aloud techniques (Pressley & Afflerbach, 1995). As a further point, we only examined the strategies that students reported using, not the quality or appropriateness with which these were deployed. That is, we did not measure the quality of students' evidence comprehension; rather, we only captured students' efforts at achieving comprehension. As such, we have limited understanding of whether students comprehended or evaluated evidence appropriately. Further research should consider students' comprehension effectiveness when encountering text-based evidence and examine which strategies students use when evaluating evidence of varying quality (e.g., strong or weak: Cacioppo, Petty, & Morris, 1983; McCrudden, Barnes, McTigue, Welch, & MacDonald, 2017; anecdotal or statistical: List & Rubenstein, 2019) and how they use such evidence in writing.

A final limitation is the lack of an association between self-reports of strategy use and the strategies reported via the typed strategy elicitation protocol. We consider this lack of an association to arise for a number of reasons. For one, our sample size was rather limited, with students reporting relatively small numbers of strategies during processing. For another, self-report measures have widely been critiqued for their lack of association with behavioral metrics (Bråten, Brante, & Strømsø, 2018; List & Alexander, 2018). Finally, Anmarkrud and Bråten's (2009) strategy measure used in this study only identified strategies as deep or surface level in nature. These general categories may not have directly mapped onto the more fine-grained strategy categories identified in the typed strategy reports captured in our study.

Second, future work should consider the role of other individual difference factors, namely, epistemic beliefs and reading ability as impacting students' reasoning about text-based evidence. Indeed, both epistemic beliefs (Bråten, Britt, Strømsø, & Rouet, 2011) and reading ability (Bråten & Strømsø, 2009; Strømsø, Bråten, & Samuelstuen, 2008) have been found to be associated with more general measures of reasoning about multiple texts (e.g., multiple-text comprehension). Likewise, prior work has found previous writing achievement (Pajares & Valiante, 1999) and writing instruction (e.g., Gil et al., 2010; Wiley & Voss, 1999) to affect students' writing performance. Nevertheless, students' baseline writing skills or writing backgrounds were not assessed in this study. Future studies should consider examining how these individual difference factors may influence students' attention to and reconciliation of evidence, as reflected in their writing. Additionally, future work should consider examining students' evidence-based reasoning across various text genres (Wolfe & Mienko, 2007). To the extent that expository vis-à-vis more narrative texts have been found to be more difficult for students to comprehend, presenting text-based evidence across genres represents an important direction for future work. Indeed, reading skills may interact with texts' genre in determining the extent of students' reasoning about text-based evidence (Ozuru, Best, & McNamara, 2004). Further, text genres (e.g., narrative vs. scientific) and students' reading skills may also interact with the mode of think-aloud used, in terms of their effects on students' strategy reports (Muñoz et al., 2006).

Third, participants in the present study were mostly White females. Prior research (Bråten & Strømsø, 2006) has shown that female undergraduate students tend to report more strategy use than male students while reading. Likewise,

although racial and ethnic differences in evidence evaluation have been studied to a limited extent, Strømsø, Bråten, Anmarkrud, and Ferguson (2016) found differences in students' beliefs regarding source trustworthiness between ethnic-minority and ethnic-majority students in Norway. Appiah (2003; see also Knobloch-Westerwick, Appiah, & Alter, 2008) has suggested that there may be some differences in the information use behaviors of Black and White individuals in the United States, particularly as mediated by Black information users' ethnic identity (Appiah, 2004). Still, the role of race/ethnicity and ethnic identification has not been examined within the context of evidence evaluation, per se. Future research should recruit a more diverse sample to examine whether there are differences in evidence-related processing strategies among students from different demographic groups. Moreover, although we used prior knowledge as a covariate in analyses, students in our sample had comparatively low prior knowledge. Further work should consider the nature of evidence-based reasoning in more knowledgeable samples. This is particularly the case given the role of domain knowledge in evidence evaluation (Arocha, Patel, & Patel, 1993; Hauslein, Good, & Cummins, 1992; Rouet et al., 1996; Rouet, Favart, Britt, & Perfetti, 1997; Wineburg, 1991).

Fourth, we did not explicitly ask students to evaluate text-based evidence prior to reading, nor did we explicitly ask them to attend to the different interpretations of the common evidence presented across texts. This may have limited the extent of students' evidence evaluation and task performance. Prior work has found that students tend to report more evaluations when explicitly asked to do so (Gerjets et al., 2011). Likewise, providing students with explicit task instructions to counterargue (Nussbaum, Kardash, & Graham, 2005; Nussbaum & Schraw, 2007) or other supports (e.g., argument diagrams representing two-sided arguments) has been found to improve argumentative writing (Nesbit, Niu, & Liu, 2019; van Amelsvoort & Schilperoord, 2018). Therefore, future work exploring links between strategy use and writing quality should consider explicitly directing students to evaluate text-based evidence during reading and to address evidence from opposing sides in their argumentative writing. At the same time, students in this study demonstrated fairly limited prior knowledge, and the texts we constructed for this study, although based on existing evidence on the construction of Stonehenge and the Pyramid of the Sun, were created using selective evidence and described in a deliberately parallel fashion across texts. Both of these features may have limited students' abilities to engage in evidence evaluation during processing.

Conclusions

In this study, we examined students' evidence-related processing strategies during reading, their ultimate evidence use in writing, and the association between these. Findings indicate that although students most commonly reported using restating as a strategy while processing text-based evidence, they also reported a considerable amount of evidence evaluation. Nevertheless, students performed poorly on measures reflecting both the quantity and quality of evidence use in their writing. This limited performance points to the need to provide students with guidelines on how much evidence is needed for a writing task and to prompt students to attend to conflicting evidence while reading multiple texts. At the same time, we found both surface-level (i.e., restating) and deep-level (i.e., evaluating, elaborating) strategies and personal reactions to evidence to be associated with the quality of students' common evidence use in writing. This suggests a possible direction for future intervention work, namely, to test whether teaching students evidence-related processing strategies could improve their use of text-based evidence in writing.

REFERENCES

Afflerbach, P., & Cho, B.-Y. (2009). Identifying and describing constructively responsive comprehension strategies in new and traditional forms of reading. In S.E. Israel & G.G. Duffy (Eds.), *Handbook of research on reading comprehension* (pp. 69–90). New York, NY: Routledge.

Anmarkrud, Ø., & Bråten, I. (2009). Motivation for reading comprehension. *Learning and Individual Differences*, *19*(2), 252–256. https://doi.org/10.1016/j.lindif.2008.09.002

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. https://doi.org/10.1016/j.lindif.2013.01.007

Anmarkrud, Ø., McCrudden, M.T., Bråten, I., & Strømsø, H.I. (2013). Task-oriented reading of multiple documents: Online comprehension processes and offline products. *Instructional Science*, *41*, 873–894. https://doi.org/10.1007/s11251-013-9263-8

Appiah, O. (2003). Americans online: Differences in surfing and evaluating race-targeted web site: By black and white users. *Journal of Broadcasting & Electronic Media*, *47*(4), 537–555. https://doi.org/10.1207/s15506878jobem4704_4

Appiah, O. (2004). Effects of ethnic identification on web browsers' attitudes toward and navigational patterns on race-targeted sites. *Communication Research*, *31*(3), 312–337. https://doi.org/10.1177/0093650203261515

Applebee, A.N. (1984). Writing and reasoning. *Review of Educational Research*, *54*(4), 577–596. https://doi.org/10.3102/00346543054004577

Arocha, J.F., Patel, V.L., & Patel, Y.C. (1993). Hypothesis generation and the coordination of theory and evidence in novice diagnostic reasoning. *Medical Decision Making*, *13*(3), 198–211. https://doi.org/10.1177/0272989X9301300305

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*, 737–766. https://doi.org/10.1007/s11251-015-9359-4

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. https://doi.org/10.3758/s13421-011-0160-6

Bråten, I., Brante, E.W., & Strømsø, H.I. (2018). What really matters: The role of behavioural engagement in multiple document literacy tasks. *Journal of Research in Reading*, *41*(4), 680–699. https://doi.org/10.1111/1467-9817.12247

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. https://doi.org/10.1080/00461520.2011.538647

Bråten, I., McCrudden, M.T., Stang Lund, E., Brante, E.W., & Strømsø, H.I. (2018). Task-oriented learning with multiple documents: Effects of topic familiarity, author expertise, and content relevance on document selection, processing, and use. *Reading Research Quarterly*, *53*(3), 345–365. https://doi.org/10.1002/rrq.197

Bråten, I., & Strømsø, H.I. (2003). A longitudinal think-aloud study of spontaneous strategic processing during the reading of multiple expository texts. *Reading and Writing*, *16*, 195–218. https://doi.org/10.1023/A:1022895207490

Bråten, I., & Strømsø, H.I. (2006). Epistemological beliefs, interest, and gender as predictors of internet-based learning activities. *Computers in Human Behavior*, *22*(6), 1027–1042. https://doi.org/10.1016/j.chb.2004.03.026

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. https://doi.org/10.1080/01638530902959646

Bråten, I., & Strømsø, H.I. (2011). Measuring strategic processing when students read multiple texts. *Metacognition and Learning*, *6*, 111–130. https://doi.org/10.1007/s11409-011-9075-7

Bråten, I., Strømsø, H.I., & Britt, M.A. (2009). Trust matters: Examining the role of source evaluation in students' construction of meaning within and across multiple texts. *Reading Research Quarterly*, *44*(1), 6–28. https://doi.org/10.1598/RRQ.44.1.1

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

Britt, M.A., & Sommer, J. (2004). Facilitating textual integration with macro-structure focusing tasks. *Reading Psychology*, *25*(4), 313–339. https://doi.org/10.1080/02702710490522658

Cacioppo, J.T., Petty, R.E., & Morris, K.J. (1983). Effects of need for cognition on message evaluation, recall, and persuasion. *Journal of Personality and Social Psychology*, *45*(4), 805–818. https://doi.org/10.1037/0022-3514.45.4.805

Cerdán, R., & Vidal-Abarca, E. (2008). The effects of tasks on integrating information from multiple documents. *Journal of Educational Psychology*, *100*(1), 209–222. https://doi.org/10.1037/0022-0663.100.1.209

Chinn, C.A., & Brewer, W.F. (2001). Models of data: A theory of how people evaluate data. *Cognition and Instruction*, *19*(3), 323–393. https://doi.org/10.1207/S1532690XCI1903_3

Cho, B.-Y., Afflerbach, P., & Han, H. (2018). Strategic processing in accessing, comprehending, and using multiple sources online. In J.L.G. Braasch, I. Bråten, & M.T. McCrudden (Eds.), *Handbook of multiple source use* (pp. 133–150). New York, NY: Routledge.

College Board. (n.d.). *K–12 educator brief: The college and career readiness benchmarks for the SAT® suite of assessments*. Retrieved from https://collegereadiness.collegeboard.org/pdf/educator-benchmark-brief.pdf

Cromley, J.G., Snyder-Hogan, L.E., & Luciw-Dubas, U.A. (2010). Cognitive activities in complex science text and diagrams. *Contemporary Educational Psychology*, *35*(1), 59–74. https://doi.org/10.1016/j.cedpsych.2009.10.002

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

De La Paz, S., Ferretti, R., Wissinger, D., Yee, L., & MacArthur, C. (2012). Adolescents' disciplinary use of evidence, argumentative strategies, and organizational structure in writing about historical controversies. *Written Communication*, *29*(4), 412–454. https://doi.org/10.1177/0741088312461591

Dinsmore, D.L., & Alexander, P.A. (2012). A critical discussion of deep and surface processing: What it means, how it is measured, the role of context, and model specification. *Educational Psychology Review*, *24*, 499–567. https://doi.org/10.1007/s10648-012-9198-7

Dinsmore, D.L., Hattan, C., & List, A. (2017). A meta-analysis of strategy use and performance in the model of domain learning. In H. Five & D.L. Dinsmore (Eds.), *The model of domain learning* (pp. 37–55). New York, NY: Routledge.

Du, H., & List, A. (2020). Researching and writing based on multiple texts. *Learning and Instruction*, *66*, article 101297. https://doi.org/10.1016/j.learninstruc.2019.101297

Duncan, R.G., Chinn, C.A., & Barzilai, S. (2018). Grasp of evidence: Problematizing and expanding the Next Generation Science Standards' conceptualization of evidence. *Journal of Research in Science Teaching*, *55*(7), 907–937. https://doi.org/10.1002/tea.21468

Flower, L., Stein, V., Ackerman, J., Kantz, M.J., McCormick, K., & Peck, W.C. (1990). *Reading to write: Exploring a cognitive and social process*. New York, NY: Oxford University Press.

Gerjets, P., Kammerer, Y., & Werner, B. (2011). Measuring spontaneous and instructed evaluation processes during web search: Integrating concurrent thinking-aloud protocols and eye-tracking data. *Learning and Instruction*, *21*(2), 220–231. https://doi.org/10.1016/j.learninstruc.2010.02.005

Gil, L., Bråten, I., Vidal-Abarca, E., & Strømsø, H.I. (2010). Summary versus argument tasks when working with multiple documents: Which is better for whom? *Contemporary Educational Psychology*, *35*(3), 157–173. https://doi.org/10.1016/j.cedpsych.2009.11.002

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. https://doi.org/10.1002/RRQ.027

Hagen, Å.M., Braasch, J.L., & Bråten, I. (2014). Relationships between spontaneous note-taking, self-reported strategies and comprehension when reading multiple texts in different task conditions. *Journal of Research in Reading*, *37*(S1), S141–S157. https://doi.org/10.1111/j.1467-9817.2012.01536.x

Hauslein, P.L., Good, R.G., & Cummins, C.L. (1992). Biology content cognitive structure: From science student to science teacher. *Journal of Research in Science Teaching*, *29*(9), 939–964. https://doi.org/10.1002/tea.3660290905

Hoeken, H., & Hustinx, L.G.M.M. (2003). The relative persuasiveness of anecdotal, statistical, causal, and expert evidence. In F.H. van Eemeren, J.A. Blair, & C.A. Willard (Eds.), *Proceedings of the fifth Conference of the International Society for the Study of Argumentation* (pp. 497–502). Amsterdam, Netherlands: Sic Stat.

Iordanou, K., & Constantinou, C.P. (2014). Developing pre-service teachers' evidence-based argumentation skills on socio-scientific issues. *Learning and Instruction*, *34*, 42–57. https://doi.org/10.1016/j.learninstruc.2014.07.004

Kammerer, Y., & Gerjets, P. (2012). Effects of search interface and internet-specific epistemic beliefs on source evaluations during web search for medical information: An eye-tracking study. *Behaviour & Information Technology*, *31*(1), 83–97. https://doi.org/10.1080/0144929X.2011.599040

Kammerer, Y., Meier, N., & Stahl, E. (2016). Fostering secondary-school students' intertext model formation when reading a set of websites: The effectiveness of source prompts. *Computers & Education*, *102*, 52–64. https://doi.org/10.1016/j.compedu.2016.07.001

Kienhues, D., Stadtler, M., & Bromme, R. (2011). Dealing with conflicting or consistent medical information on the web: When expert information breeds laypersons' doubts about experts. *Learning and Instruction*, *21*(2), 193–204. https://doi.org/10.1016/j.learninstruc.2010.02.004

Klein, P.D. (1999). Reopening inquiry into cognitive processes in writing-to-learn. *Educational Psychology Review*, *11*, 203–270. https://doi.org/10.1023/A:1021913217147

Knobloch-Westerwick, S., Appiah, O., & Alter, S. (2008). News selection patterns as a function of race: The discerning minority and the indiscriminating majority. *Media Psychology*, *11*(3), 400–417. https://doi.org/10.1080/15213260802178542

Kobayashi, K. (2014). Students' consideration of source information during the reading of multiple texts and its effect on intertextual conflict resolution. *Instructional Science*, *42*, 183–205. https://doi.org/10.1007/s11251-013-9276-3

List, A. (2020). Six questions regarding strategy use when learning from multiple texts. In D.L. Dinsmore, L.K. Fryer, & M.M. Parkinson (Eds.), *Handbook of strategies and strategic processing* (pp. 119–140). New York, NY: Routledge.

List, A., & Alexander, P.A. (2017). Cognitive affective engagement model of multiple source use. *Educational Psychologist*, *52*(3), 182–199. https://doi.org/10.1080/00461520.2017.1329014

List, A., & Alexander, P.A. (2018). Corroborating students' self-reports of source evaluation. *Behaviour & Information Technology*, *37*(3), 198–216. https://doi.org/10.1080/0144929X.2018.1430849

List, A., Alexander, P.A., & Stephens, L.A. (2017). Trust but verify: Examining the association between students' sourcing behaviors and ratings of text trustworthiness. *Discourse Processes*, *54*(2), 83–104. https://doi.org/10.1080/0163853X.2016.1174654

List, A., Du, H., & Lee, H.Y. (2020). How do students integrate multiple texts? An investigation of top-down processing. *European Journal of Psychology of Education*. Advance online publication. https://doi.org/10.1007/s10212-020-00497-y

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. https://doi.org/10.1016/j.cedpsych.2019.03.003

List, A., & Rubenstein, L.D. (2019). Understanding susceptibility to educational inaccuracies: Examining the likelihood of adoption model. In P. Kendeou, D.H. Robinson, & M.T. McCrudden (Eds.), *Misinformation and fake news in education*, (pp. 29–54). Charlotte, NC: Information Age.

Lombardi, D., Sinatra, G.M., & Nussbaum, E.M. (2013). Plausibility reappraisals and shifts in middle school students' climate change conceptions. *Learning and Instruction*, *27*, 50–62. https://doi.org/10.1016/j.learninstruc.2013.03.001

Mason, L., Ariasi, N., & Boldrin, A. (2011). Epistemic beliefs in action: Spontaneous reflections about knowledge and knowing during online information searching and their influence on learning. *Learning and Instruction*, *21*(1), 137–151. https://doi.org/10.1016/j.learninstruc.2010.01.001

Mateos, M., Rijlaarsdam, G., Martín, E., Cuevas, I., Van den Bergh, H., & Solari, M. (2020). Learning paths in synthesis writing: Which learning path contributes most to which learning outcome? *Instructional Science*, *48*, 137–157. https://doi.org/10.1007/s11251-020-09508-3

McCarthy, P.M., Guess, R.H., & McNamara, D.S. (2009). The components of paraphrase evaluations. *Behavior Research Methods*, *41*, 682–690. https://doi.org/10.3758/BRM.41.3.682

McCrudden, M.T., Barnes, A., McTigue, E.M., Welch, C., & MacDonald, E. (2017). The effect of perspective-taking on reasoning about strong and weak belief-relevant arguments. *Thinking & Reasoning*, *23*(2), 115–133. https://doi.org/10.1080/13546783.2016.1234411

McNamara, D.S. (2004). SERT: Self-explanation reading training. *Discourse Processes*, *38*(1), 1–30. https://doi.org/10.1207/s15326950dp3801_1

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

Monte-Sano, C., & De La Paz, S. (2012). Using writing tasks to elicit adolescents' historical reasoning. *Journal of Literacy Research*, *44*(3), 273–299. https://doi.org/10.1177/1086296X12450445

Muñoz, B., Magliano, J.P., Sheridan, R., & McNamara, D.S. (2006). Typing versus thinking aloud when reading: Implications for computer-based assessment and training tools. *Behavior Research Methods*, *38*, 211–217. https://doi.org/10.3758/BF03192771

Nesbit, J., Niu, H., & Liu, Q. (2019). Cognitive tools for scaffolding argumentation. In O.O. Adesope & A.G. Rud (Eds.), *Contemporary technologies in education: Maximizing student engagement, motivation, and learning* (pp. 97–117). Cham, Switzerland: Palgrave Macmillan.

Novak, J.D. (2010). *Learning, creating, and using knowledge: Concept maps as facilitative tools in schools and corporations*. New York, NY: Routledge.

Nussbaum, E.M., Kardash, C.M., & Graham, S.E. (2005). The effects of goal instructions and text on the generation of counterarguments during writing. *Journal of Educational Psychology*, *97*(2), 157–169. https://doi.org/10.1037/0022-0663.97.2.157

Nussbaum, E.M., & Schraw, G. (2007). Promoting argument-counterargument integration in students' writing. *Journal of Experimental Education*, *76*(1), 59–92. https://doi.org/10.3200/JEXE.76.1.59-92

Olinghouse, N.G., & Graham, S. (2009). The relationship between the discourse knowledge and the writing performance of elementary-grade students. *Journal of Educational Psychology*, *101*(1), 37–50. https://doi.org/10.1037/a0013462

Ozuru, Y., Best, R., & McNamara, D.S. (2004). Contribution of reading skill to learning from expository texts. In K. Forbus, D. Gentner, & T. Regier (Eds.), *Proceedings of the 26th annual meeting of the Cognitive Science Society* (pp. 1071–1076). Mahwah, NJ: Erlbaum.

Pajares, F., & Valiante, G. (1999). Grade level and gender differences in the writing self-beliefs of middle school students. *Contemporary Educational Psychology*, *24*(4), 390–405. https://doi.org/10.1006/ceps.1998.0995

Perfetti, C.A., Rouet, J.-F., & Britt, M.A. (1999). Toward a theory of documents representation. In H. van Oostendorp & S.R. Goldman (Eds.), *The construction of mental representations during reading* (pp. 88–108). Mahwah, NJ: Erlbaum.

Pressley, M., & Afflerbach, P. (1995). *Verbal protocols in reading: The nature of constructively responsive reading*. Hillsdale, NJ: Erlbaum.

Reznitskaya, A., Kuo, L.J., Glina, M., & Anderson, R.C. (2009). Measuring argumentative reasoning: What's behind the numbers? *Learning and Individual Differences*, *19*(2), 219–224. https://doi.org/10.1016/j.lindif.2008.11.001

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

Rouet, J.-F., Favart, M., Britt, M.A., & Perfetti, C.A. (1997). Studying and using multiple documents in history: Effects of discipline expertise. *Cognition and Instruction*, *15*(1), 85–106. https://doi.org/10.1207/s1532690xci1501_3

Sandoval, W.A., & Millwood, K.A. (2005). The quality of students' use of evidence in written scientific explanations. *Cognition and Instruction*, *23*(1), 23–55. https://doi.org/10.1207/s1532690xci2301_2

Scardamalia, M., & Paris, P. (1985). The function of explicit discourse knowledge in the development of text representations and composing strategies. *Cognition and Instruction*, *2*(1), 1–39. https://doi.org/10.1207/s1532690xci0201_1

Spivey, N.N. (1990). Transforming texts: Constructive processes in reading and writing. *Written Communication*, *7*(2), 256–287. https://doi.org/10.1177/0741088390007002004

Stadtler, M., & Bromme, R. (2007). Dealing with multiple documents on the WWW: The role of metacognition in the formation of documents models. *International Journal of Computer-Supported Collaborative Learning*, *2*, 191–210. https://doi.org/10.1007/s11412-007-9015-3

Stadtler, M., & Bromme, R. (2014). The content–source integration model: A taxonomic description of how readers comprehend conflicting scientific information. In D.N. Rapp & J.L.G. Braasch (Eds.), *Processing inaccurate information: Theoretical and applied perspectives from cognitive science and the educational sciences* (pp. 379–402). Cambridge, MA: MIT Press.

Strømsø, H.I., & Bråten, I. (2014). Students' sourcing while reading and writing from multiple web documents. *Nordic Journal of Digital Literacy*, *9*, 92–111.

Strømsø, H.I., Bråten, I., Anmarkrud, Ø., & Ferguson, L.E. (2016). Beliefs about justification for knowing when ethnic majority and ethnic minority students read multiple conflicting documents. *Educational Psychology*, *36*(4), 638–657. https://doi.org/10.1080/01443410.2014.920080

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. https://doi.org/10.1080/07370008.2013.769994

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. https://doi.org/10.1016/j.learninstruc.2007.11.001

Thomm, E., & Bromme, R. (2016). How source information shapes lay interpretations of science conflicts: Interplay between sourcing, conflict explanation, source evaluation, and claim evaluation. *Reading and Writing*, *29*, 1629–1652. https://doi.org/10.1007/s11145-016-9638-8

Toulmin, S. (1958). *The use of argument*. Cambridge, UK: Cambridge University Press.

van Amelsvoort, M., & Schilperoord, J. (2018). How number and size of text boxes in argument diagrams affect opinions. *Learning and Instruction*, *57*, 57–70. https://doi.org/10.1016/j.learninstruc.2018.03.003

van Gog, T., Paas, F., van Merriënboer, J.J., & Witte, P. (2005). Uncovering the problem-solving process: Cued retrospective reporting versus concurrent and retrospective reporting. *Journal of Experimental Psychology: Applied*, *11*(4), 237–244. https://doi.org/10.1037/1076-898X.11.4.237

Walraven, A., Brand-Gruwel, S., & Boshuizen, H.P. (2010). Fostering transfer of websearchers' evaluation skills: A field test of two transfer theories. *Computers in Human Behavior*, *26*(4), 716–728. https://doi.org/10.1016/j.chb.2010.01.008

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

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

Wolfe, M.B., & Goldman, S.R. (2005). Relations between adolescents' text processing and reasoning. *Cognition and Instruction*, *23*(4), 467–502. https://doi.org/10.1207/s1532690xci2304_2

Wolfe, M.B., & Mienko, J.A. (2007). Learning and memory of factual content from narrative and expository text. *British Journal of Educational Psychology*, *77*(3), 541–564. https://doi.org/10.1348/000709906X143902

Submitted October 11, 2019
Final revision received June 22, 2020
Accepted June 23, 2020

HONGCUI DU (corresponding author) is a doctoral student in the Department of Educational Psychology, Counseling, and Special Education at The Pennsylvania State University, State College, USA; email hzd27@psu.edu.

ALEXANDRA LIST is an assistant professor in the Department of Educational Psychology, Counseling, and Special Education at The Pennsylvania State University, State College, USA; email azl261@psu.edu. Her work focuses on examining students' learning from multiple texts and the higher order cognitive processes (e.g., evaluation, integration) involved.

APPENDIX

Correlations Between Self-Report Processing Strategies and Other Processing and Performance Variables

	Evidence processing						Performance	
Self-reported processing strategies	Restating	Elaborative question	Evaluation	Elaboration	Personal reaction	Self-regulation	Total evidence use	Common evidence use
Mean (standard deviation)	3.13 (3.63)	1.51 (1.99)	1.81 (2.04)	1.72 (1.95)	0.53 (0.98)	0.40 (0.69)	1.96 (1.39)	1.44 (1.31)
Range	10	7	8	9	5	3	5	4
Deeper strategies	−.21	.21	.11	.16	.19	.12	.12	−.02
Surface-level strategies	.32**	−.05	−.16	−.11	−.08	−.07	.20	.02