



## On the basis of source: Impacts of individual differences on multiple-document integrated reading and writing tasks

Kathryn S. McCarthy<sup>a,\*</sup>, Eleanor F. Yan<sup>a</sup>, Laura K. Allen<sup>b</sup>, Allison N. Sonia<sup>b</sup>, Joseph P. Magliano<sup>a</sup>, Danielle S. McNamara<sup>c</sup>

<sup>a</sup> Georgia State University, USA

<sup>b</sup> University of New Hampshire, USA

<sup>c</sup> Arizona State University, USA

### ARTICLE INFO

#### Keywords:

Multiple-documents inquiry  
Source-based writing  
Individual differences  
Prior knowledge  
Reading skill  
Writing skill

### ABSTRACT

Few studies have explored how general skills in both reading and writing influence performance on integrated, source-based writing. The goal of the present study was to consider the relative contributions of reading and writing ability on multiple-document integrative reading and writing tasks. Students in the U.S. ( $n = 94$ ) completed two tasks in which they read text sets about a socioscientific issue, generated constructed responses while reading, and then composed integrated essays. They also completed individual difference measures (general knowledge, reading skill, reading strategy use) and wrote independent essays to assess their writing ability. Mixed effect models revealed that general knowledge and reading skills contributed to integrated essay performance, but that once general writing ability was entered into the model, it became the strongest predictor of integrated writing scores. These results suggest the need for deeper consideration of the role of writing skills in integrated reading and writing tasks.

### 1. Introduction

A variety of academic and professional activities rely on *source-based* reading and writing. Source-based tasks involve writing about information gathered from an existing source or set of sources. For example, an elementary school student may write a book report about a novel or a Supreme Court justice might draw upon old court documents and laws to compose a dissenting opinion. In the current study, we are specifically interested in source-based writing tasks. In these tasks, students are asked to read multiple documents from varying sources and to use the information in the documents to compose an essay intended to summarize, describe, inform, or persuade. These multiple-document integrated reading and writing tasks (henceforth, integrated writing tasks) have become increasingly common. This is, in part, because of the increased availability of (often conflicting) information on the internet which has influenced the ways in which we encounter and use information both in and out of the classroom (Magliano et al., 2018).

In the classroom, integrated writing tasks are sometimes used to provide writing instruction or assess writing skills. They are also used in classrooms and research studies as a means to assess learning outcomes,

often to approximate the extent to which a student has understood the topic or phenomena. In the latter case, integrated writing tasks are intended as a measure of comprehension and learning, and the extent to which writing skills play a role is often ignored. This is problematic, given that source-based integrated writing tasks rely on both lower-level and higher-order literacy skills related to comprehension as well as a number of skills related to production including composition and argumentation (e.g., List & Alexander, 2019; Spivey & King, 1989).

Understanding how various literacy skills and knowledge come together in an integrated writing task is necessary to provide more targeted support for students. While there is a growing body of research examining multiple-document literacies (see Braasch et al., 2018; Van Meter et al., 2020), multiple gaps in the literature remain. These gaps point to a need to devote greater consideration toward understanding the potential roles of individual differences in these integrated writing tasks. We examine these issues below.

#### 1.1. Theories of comprehension

In the majority of research on multiple-document comprehension,

\* Corresponding author. Department of Learning Sciences, Georgia State University, P.O. Box 3978, Atlanta, GA, 30302, USA.

E-mail address: [kmccarthy12@gsu.edu](mailto:kmccarthy12@gsu.edu) (K.S. McCarthy).

the researchers' objective is to assess the degree to which the student understands and integrates information from the documents (Goldman & Scardamalia, 2013; Perfetti et al., 1999; Wiley et al., 2017). Usually, the students' understanding is assessed via an integrated writing task (Primor & Katzir, 2018). As such, the majority of the research and methodologies are guided by assumptions inherent to models and theories of how individuals comprehend discourse. Discourse comprehension theories posit that readers construct a *mental model* of text information as they read. A *coherent* mental model is one that accurately represents the information that was explicitly conveyed, inferences generated, and the relations between these sources of information (McNamara & Magliano, 2009). For example, the most prevalent theory of discourse comprehension, the Construction-Integration model, specifies that readers construct a multi-level mental model during reading (Kintsch, 1988). The *surface code* contains the exact words and sentence structures from the text, whereas the *textbase* reflects a more abstracted gist of the text information. Finally, the *situation model* is composed of the inferences that connect different ideas from across the text as well as relevant information from the reader's prior knowledge. Thus, a mental model that is more elaborated and well-integrated is likely to afford deeper understanding and retention.

Much of this early work in discourse comprehension was aimed at explaining how a reader understood a single text. That is, many of these studies involve participants reading a text and answering comprehension questions that probe for the structure of the reader's mental model (see McCarthy et al., 2018). More recently, however, researchers have acknowledged that these "text-then-test" reading tasks do not reflect the types of comprehension tasks that learners engage in throughout their daily lives (Magliano et al., 2018; Sabatini et al., 2019). Thanks in large part to the internet and growing demands for more complex reasoning skills, successful learners are often required to engage in multiple-document reading and writing tasks (Bråten & Braasch, 2017; Britt et al., 2018, 2012; OECD, 2019). In these tasks, students still need to form connections within a text (i.e., intra-textual inferences), but they must also generate *inter-textual inferences* that connect information across texts (Salmerón et al., 2010). Often these texts might be written for different purposes and may also include information that is not only complementary, but also discrepant, contradictory, or otherwise conflicting (Braasch & Bråten, 2017; Scharrer & Salmerón, 2016). As a result, readers often need to maintain an idea of "who said what" along with their general representations of the texts. Thus, theories of multiple-document comprehension do not aim to refute single text comprehension models, but rather to extend them to consider additional factors relevant to how we represent and integrate multiple documents while reading.

One of the most influential models of multiple-document comprehension is the *Documents Model Framework* (DMF; Britt et al., 1999, 2012; Perfetti et al., 1999). The DMF extends the description of mental model building through the addition of both an *integrated mental model* of the global situation described in multiple texts and an *intertext model* which relies on the reader's attention to source features (e.g., author, date of publication, publication type). More recently, the RESOLV model (Britt et al., 2018; Rouet et al., 2017) has been developed to expand on the DMF to provide a more general explanation of goal-driven reading. In addition to the integrated mental model and the intertext model described in the DMF, the RESOLV model includes two additional representations. The *context model* includes a representation of the current reading situation and task requirements while the *task model* includes a representation of the reader's goals and plans. As a reader progresses through a multiple-document inquiry task, these models help the learner to monitor the completeness of their mental representation and to evaluate the quality of their comprehension in light of task goals.

Both single and multiple-document comprehension theories highlight the importance of individual differences. Individual differences such as prior knowledge, reading skill (e.g., Ozuru et al., 2009), and strategy use (e.g., Anmarkrud et al., 2013) are strong predictors of

comprehension performance. Research on prior knowledge shows that more knowledgeable readers can more quickly access and organize information and are better able to generate relevant inferences that support the development of a coherent mental model (see Dochy et al., 1999). Research has also shown that general reading skills, as measured by standardized measures such as the Gates-MacGinitie Reading Test and the Nelson-Denny are predictive of performance on content-based, task-oriented reading (e.g., learning from a text; Ozuru et al., 2009). More successful readers tend to engage in more effective metacognitive and comprehension strategies (e.g., Anmarkrud et al., 2013; Cote et al., 1998; Cromley & Azevedo, 2007; McNamara, 2007). Interestingly, explicit knowledge and use of metacognitive strategies have been shown to be particularly important for successful multiple-documents comprehension (Karimi, 2015).

Related to strategy use, a number of single and multiple-document comprehension studies have explored the extent to which prompting students to use different comprehension strategies can impact how they learn from text. Research from think-aloud studies (e.g., Coté et al., 1998; Wineburg, 1991; Wolfe & Goldman, 2005) indicate that more skilled readers tend to engage in more active and integrative strategies such as *self-explanation* and *source evaluation*. Further, intervention work has shown that supporting students' use of these strategies can lead to better comprehension, especially for less skilled and less knowledgeable readers (e.g., Britt & Aglinskas, 2002; McNamara et al., 2004). In the current work, we provide a brief constructed response instructional prompt to bias the readers toward using these strategies. While such prompts are less effective than longer-term interventions, we can use this manipulation to further explore the role of strategy use on the quality of integrated writing.

Research in multiple-document comprehension often relies on a written product to assess the extent to which readers have successfully processed, integrated, and comprehended the documents. After reading, participants are asked to write a summary, description, or argument and the written product is evaluated for evidence of source use (referencing the source information of the document), content coverage, and the degree of integration. Integration is measured in a number of ways. A review of multiple-document integration tasks (Primor & Katzir, 2018) revealed that the majority of studies evaluate students' written products for the degree of content coverage, source use, and the extent to which participants address conflicting viewpoints. Few, if any, of these studies examine aspects of writing such as mechanics or style.

This is, in large part because this area of work tends to use the essay as a window into comprehension, rather than as a task unto itself (see also Wiley et al., 2018). As a result, these studies have yielded a number of insights into the complex processes involved in multiple-document comprehension, but have far less to contribute to an understanding of integrated writing. For example, the Integrated Framework of Multiple Texts (IF-MT; List & Alexander, 2019) outlines three stages: preparation, execution, and production. The execution stage includes identification of the behavioral (e.g., search, inter- and intra-textual navigation), metacognitive (e.g., monitoring), and cognitive skills and strategies (e.g., knowledge activation, integration, perspective taking) involved in multiple-document processing. By contrast, the production stage is far less developed. The framework identifies the *types* of products (e.g., arguments, opinions, report), but does not explore or identify the skills and strategies involved in the writing process nor the extent to which these processes are similar or different from the comprehension process.

Generally speaking, the essays in these studies are not explicitly scored for the quality of writing, but rather on the degree of key content coverage (e.g., Wiley et al., 2009, 2014), the number of sources referenced (e.g., Bråten et al., 2014; Britt & Aglinskas, 2002), the degree of integration (e.g., List et al., 2019) and organization (e.g., Wiley & Voss, 1996, 1999), or the quality of argumentation (e.g., Anmarkrud et al., 2013; Bråten et al., 2014). Even though writing quality is not the explicit focus of these measures, it is likely that students' ability to represent this information is, at least in part, dependent on their writing skills. Thus, a

limitation in the current body of work is that performance on integrated essays is often assumed to be a measure of comprehension processes; however, there is little recognition that differences in comprehension may be obscured by differences in production.

## 1.2. Theories of writing

Theories of discourse *production* exist largely separately from theories of discourse *comprehension*, although both processes require multiple strategies and sources of knowledge to complete. At its simplest, discourse production in the form of *writing* requires an individual to transcribe their thoughts into a written form with minimal spelling and mechanical errors. High-quality writing, however, is more than simply a transcription of thought – indeed, successful writing transforms thought into discourse that is understandable by the intended reader (Graham, 2018). Theoretical models have delineated the writing process into three primary components: planning, translating, and reviewing (Flower & Hayes, 1981; see also Hayes, 1996, 2012). In the *planning* phase, a writer generates and organizes their ideas. These ideas are then *translated* into words on a page and *reviewed* to determine whether they have achieved the writer's goals and may subsequently be revised. Importantly, theories dictate that this is a nonlinear process; thus, a writer may frequently switch amongst these processes based on a variety of factors due to constraints of the task or the individual.

Skilled writing is not only marked by differences in content, but also in style. Skilled writers produce longer essays (Ferrari et al., 1998; Haswell, 2000; McNamara et al., 2010, 2013) with fewer mechanical errors (Ferrari et al., 1998). They also tend to use more abstract language (Crossley, Weston, et al., 2011; McNamara et al., 2010, 2013) and more complex syntax (McCutchen et al., 1994). The individual differences that relate to skilled writing encompass a wide range of features. A number of models of writing proficiency have attempted to incorporate individual differences in their models ranging from world knowledge (McCutchen, 1986; Olinghouse et al., 2015) to reading comprehension (Allen et al., 2014; Fitzgerald & Shanahan, 2000; Tierney & Shanahan, 1991). Importantly, strong writers have a greater understanding of the writing process itself. Saddler and Graham (2007), for example, found that less skilled writers showed a weaker understanding of writing goals ( $d = -1.13$ ), were less knowledgeable of the differences between high- and low-quality writing ( $d = -0.98$ ), and had less knowledge of successful writing strategies ( $d = -1.10$ ). These results are important, as they indicate that variability in knowledge of *how* to write can influence the quality of a written text.

Generally speaking, the majority of work that has informed theories of writing has focused on independent essays in which writers are responding to a prompt by drawing upon their own knowledge and experience. However, consistent with the uptick in research on multiple-documents comprehension, there has been an increase in research on source-based writing and, in particular, multiple-document writing. Research in these integrated writing tasks (also referred to as hybrid or synthesis writing; e.g., Mateos et al., 2014; Spivey, 1997) have focused on the idea that writing tasks are not linear, but rather a recursive process in which the writer needs to move between processes. This is particularly important for integrated writing because the student needs to not only move between production processes, but also between reading and writing (Spivey, 1997; Spivey & King, 1989). Indeed, more recursive writers (e.g., those who go back-and-forth between sources and the essay) tend to write higher quality essays (e.g., Martinez et al., 2015; Vandermeulen et al., 2020). In addition, integrated writing not only requires the planning, drafting, and revising identified in the Flower and Hayes (1981) model, but also selecting information, organizing that information, and connecting the information (see van Ockenburg et al., 2019). Classroom intervention work has demonstrated that providing instruction and practice for these integrated writing strategies can improve integrated essay quality (e.g., Mateos et al., 2018; van Ockenburg et al., 2021; Wissinger et al., 2021).

## 1.3. Individual differences that impact integrated writing

As described earlier, high-quality integrated writing requires the reader to develop a coherent mental model of the text content (reading) and to convey that information in a coherent and organized manner (writing). Although there has been an increase in the research on integrated writing, there remains unknowns about the unique contributions of reading and writing skills on an individual's ability to produce high quality integrated essays. A criticism of the work in multiple-documents comprehension research is the use of integrated writing as a measure of comprehension without considering the effects of the students' general writing skill (e.g., Primor & Katzir, 2018; Wiley et al., 2018). Conversely, many studies collect performance on a standardized reading skill test (e.g., Gates-MacGinitie, Nelson Denny) as a control variable, but few have explored the effects of general reading skill(s) on integrated writing. One such study found that skilled readers showed improvement in all three integrated writing strategies, whereas less skilled readers were only able to improve in their ability to select (but not organize or integrate) relevant information in their post-intervention essays (e.g., De La Paz & Felton, 2010). These findings suggest that reading skill impacts writing quality, but it is unclear if these effects are related to issues in comprehension, production, or both.

It is also important to consider that fundamental writing skills are crucial for producing any written product. For example, Du and List (2020) analyzed screen recordings of students completing an integrated writing task and found that the students spent very little time planning to write or revising their essay. The authors speculate that the lack of planning and revision behaviors in their data might reflect "more general difficulties with written response composition" (Du & List, 2020, p. 18). One might imagine a student who was able to select and integrate important ideas into a coherent mental representation but struggled to convey their ideas clearly and coherently. Such a student would need different supports than a student who struggles to understand the texts. Thus, exploring the contributions of general reading and writing skill in an integrated writing task is an important endeavor for learning and instruction.

## 1.4. The current study

Integrated writing requires that the learner read and understand the source(s) and then also convey their ideas in an organized and coherent fashion. Thus, it is generally assumed that integrated essays require the coordination of literacy skills related to both comprehension and production. However, much of the research on multiple-document comprehension uses integrated writing as an outcome measure without considering how students' writing skills might impact the essay. Thus, the current study examines the following question: *To what extent does students' more general writing ability predict performance on a multiple-document source-based integrated (reading and writing) essay above and beyond factors more directly related to comprehension (i.e., general knowledge, reading skill, and metacognitive reading strategies)?*

In this study, participants were asked to read multiple documents about a topic and then compose an integrated essay. To reduce the impact of a particular prompt or text set, participants completed two separate integrated writing tasks. Consistent with extant multiple-document research (e.g., Anmarkrud et al., 2014; De La Paz & Felton, 2010), the essays were evaluated on aspects of argumentation, source use and integration, and organization. These subscales have been used as evidence of comprehension in several studies (e.g., Bråten et al., 2014; Britt & Aglinskias, 2002, Wiley et al., 2009, 2014). In addition, the essays were evaluated for their general writing quality in terms of word choice, syntax, and spelling and mechanics. We then examined the extent to which performance on these integrated essays was related to individual differences in reading and writing.

As part of a larger ongoing project, we prompted participants to generate constructed responses as they read and manipulated the

constructed response prompt instruction so that some students were biased to engage in more effective comprehension strategies (self-explanation, source evaluation) as compared to a control (thinking aloud). Although strategy prompting can benefit comprehension and learning (e.g., Allen et al., 2016; Chi et al., 1994), most students need more prolonged instruction and training to use these strategies effectively (e.g., Jackson & McNamara, 2011). Thus, we did not anticipate robust effects of this manipulation, but were interested in exploring how encouraging more effective strategy use might impact the quality of the final written product.

In contrast to the manipulation, we anticipated that participants' individual differences in knowledge would be strongly related to the quality of the integrated essays. Students completed a general academic knowledge test because the impact of general knowledge has been implicated in both reading and writing. We anticipated that students with greater general knowledge would compose higher quality integrated essays due the benefits of that knowledge in both comprehension and production. Participants also completed the Gates-MacGinitie Reading Test and the Metacognitive Awareness of Reading Strategies Inventory for Multiple Documents (MARSIMD; Karimi, 2015). These measures were conceptually-aligned with aspects of *comprehension*. More skilled readers tend to generate more inferences that support the construction of a coherent mental model and tend to draw upon more sophisticated and appropriate comprehension strategies. Both reading skill and awareness of reading strategies have been implicated in multiple-document comprehension (e.g., Barzilai & Strømsø, 2018; Karimi, 2015).

In addition, participants wrote two independent SAT-style essays as a means of evaluating their general writing ability. Thus, performance on these essays was conceptually aligned with aspects of text *production*. While it stands to reason that general writing would be related to performance on the integrated writing task, few studies examining integrated writing have also included a separate writing measure to assess general writing skills.

We predicted that higher scores on general knowledge, reading skills, and reading strategies assessments would be positively related to integrated essay quality. However, it was also predicted that general knowledge and reading skills would only partially explain variance in integrated essay performance. Prior work demonstrates that reading and writing are related, but that they rely on different processes and knowledge (Allen et al., 2014). Therefore, we also predicted that performance on the independent writing tasks would predict integrated essay quality above and beyond differences in general academic knowledge, reading skill, and comprehension strategies.

In addition to these main effects, we also explored possible interactions between these individual difference measures. Previous research has shown that prior knowledge has both direct and indirect effects on learners' comprehension processes (e.g., Cromley & Azevedo, 2007; McNamara & Magliano, 2009). Skilled writers also tend to make better use of their extant knowledge. Writing research also suggests possible interactions such that more knowledgeable students can write stronger essays and conversely, more skilled writers can better leverage the knowledge they possess (e.g., McCutchen, 1986, 2000). While there is limited research directly examining the contribution of reading skill to writing tasks, research suggests complex and potentially bi-directional development between reading and writing (e.g., Graham et al., 2018; Graham & Hebert, 2011). For simplicity, we included only the two-way interactions across each of the three individual differences variables (e.g., general knowledge by reading, general knowledge by writing, reading by writing).

## 2. Method

### 2.1. Participants

In the summer of 2019, we recruited 51 high school students. In the

fall of 2019, we recruited 45 college freshmen students who had graduated high school the previous spring. Two students (both high school) were excluded from the analysis: one did not complete all four days of the study and one student's data was lost due to technical error.

Of the remaining 94 participants, all but one student (undergraduate) completed the demographic questionnaire. The sample was predominantly female, with 76 identifying as female and 17 participants identifying as male. The sample was also predominantly Black or African American (72%;  $n = 68$ ); 10 participants identified as Asian/Pacific Islander, 10 identified as Mixed or Multiracial, 4 identified as White, and 1 identified as Native American/Alaska Native. Most of the participants ( $n = 82$ ) were native English speakers. The high school students had a mean age of 16.22 ( $SD = 0.96$ ) and the college students had a mean age of 18.34 ( $SD = 0.53$ ).

### 2.2. Design & procedure

The study was completed over 4 sessions. In Session 1, participants completed a multiple-document integrated reading and writing task in the Interactive Strategy Training for Active Reading and Thinking (iSTART; McNamara et al., 2004) interface (Watanabe et al., 2019). At the beginning of the activity, participants were told that they would be "completing a reading and writing task" using a set of documents. They were first given 4 min to skim the texts to familiarize themselves with the texts before they began the main activity. After skimming, participants were directed to the constructed response instructions in which they were told that they would be reading texts and then answering comprehension questions about the text. They were then given instructions on how to generate one of three types of constructed responses (think-aloud, self-explain, source evaluation) as they read. At target sentences throughout the texts, participants were prompted to generate constructed responses. The reading and constructed response portion of the task was untimed. After reading, participants were then provided the integrative essay instructions and given 25 min to write their essay. Notably, participants were able to access the texts and their constructed responses as they composed their essay.

In Session 2, they completed a second multiple-document integrated reading and writing task with the alternate text set. In Sessions 3 and 4, participants completed the individual difference tasks. Specifically, in Session 3, they completed a multiple-document version of the Metacognitive Awareness of Reading Strategies Inventory (MARSIMD; Karimi, 2015), a general knowledge test (O'Reilly & McNamara, 2007a, 2007b), and the Gates-MacGinitie Reading test as a measure of general reading skill. In Session 4, they wrote two SAT-style persuasive independent essays as a measure of general writing skill.

### 2.3. Materials

#### 2.3.1. Multiple-document text sets

Two text sets were adapted from previous multiple-document studies (Anmarkrud et al., 2013; Ferguson et al., 2012; Strømsø et al., 2010). The texts were translations of the original materials published in Norwegian. In addition to the text content, each text was presented with the following source content: title, author's name and credentials, publication source, and date of publication. Although these texts have been translated and slightly modified for clarity and context, the original texts were found in documents that offer authentic conflicting opinions on a socioscientific controversy. The original text sets contained eight texts that described two different positions on the topics of *global warming* (i.e., the positive and negative impacts of climate change and the extent to which climate change is a manmade phenomenon) and the *linkage between cell phones and cancer* (i.e., research that does and does not support that linkage). In the present study, we reduced the number of texts in each set to four texts. We also made minor revisions to the source information and content. Dates were updated to make the texts more contemporary, and the names and publication outlets were modified to



be more culturally relevant for our US-based sample. For example, “Chief Engineer Thor Zackarisson, Mobiloperatørens landsforening (MLF)” was changed to “Zachary Thornton, UN Telecom Association (UNTelecom)”. There were also minor modifications to the text content to increase clarity and to embed key ideas that had been removed in the reduction of the number of texts.

Both text sets followed a similar pattern. One text introduced the controversy (Is climate change manmade?; Are cell phones linked to cancer?); One text provided scientific background information (i.e., how greenhouse gases affect climate; how cell phone towers work); and then two texts presented different sides of the controversy. The texts ranged from 263 to 504 words and had a Flesch-Kincaid grade level range from 9.6 to 13.5.

### 2.3.2. Constructed response instructions

Participants were randomly assigned to one of three constructed response instruction conditions: think-aloud, self-explanation, or source evaluation. In the *think-aloud* condition, participants were asked to “report your thoughts that immediately come to mind regarding how you understand the meaning of the text”. In the *self-explanation* condition, they were asked to explain the text to themselves as they read and to “Please explain the meaning of the text, elaborating beyond your initial understanding of the text.” Finally, participants in the *source evaluation* condition were asked to “reflect on the source (i.e., author, publication date/location, audience) of the text while you read. Please report your thoughts regarding how the source impacts the meaning of the text.” Participants were prompted to type their constructed responses at pre-determined target sentences throughout the texts. Each text contained between 5 and 9 target sentences and the target sentences were consistent across the three constructed response conditions. Although typed responses are somewhat different than spoken verbal protocols, typed protocols have been shown to reflect the same types of processes as spoken protocols (Muñoz et al., 2006).

### 2.3.3. Integrated essay instructions

Participants were given 25 min to compose each essay. For the global warming text set, the prompt read “Write an essay that explains the effects of climate change for life on earth and the extent to which humans are responsible.” For the Cell Phone text set, the prompt read “Write an essay that explains the effects of cell phones on humans and the extent to which cell phone use poses health risks.” As they read, participants could access both

the texts that they read and the constructed responses that they generated during reading in a tabbed window to the right of the essay box (see Fig. 1 for interface). This window also included additional instructions that encouraged the participants to use the documents as a resource and to source and elaborate upon the ideas in the documents (Appendix A for full instructions).

### 2.3.4. Individual difference measures

**2.3.4.1. Reading skill measures.** Participants completed an adapted version of the Gates-MacGinitie Reading Test (GMRT; MacGinitie, 1989). The comprehension test includes a 48-item multiple-choice test in which participants read passages and answer questions about the passages. The vocabulary test is composed of 45 vocabulary items.

**2.3.4.2. General writing measure (independent essays).** As a measure of generalized writing ability, participants wrote two SAT-style independent, or persuasive, essays. The essay prompts were adapted from publicly released SAT exam materials. These prompts asked participants to adopt a stance with regard to a central topic, and then to defend that position via evidence, examples, and/or logical reasoning. One prompt asked participants “Do images and impressions have a positive or negative effect on people?” (*Images* prompt) and the other asked “Do people achieve more success by cooperation or by competition?” (*Competition* prompt; for full prompts see Appendix B). All prompts were designed to minimize prior knowledge demands such that participants could write from experience rather than constrained educational content or source materials. Previous work has demonstrated robust prompt effects in SAT-style independent essay tasks (e.g., Kobrin et al., 2011). To attenuate this issue, participants were asked to write two essays. The order of the essays was counterbalanced across participants. Participants had 25 min to write each essay with a brief 5–10 min break between each essay.

**2.3.4.3. General knowledge test.** Participants complete a 30-item multiple-choice test that was developed in prior work examining reading comprehension, writing skill, and strategy acquisition skill (Allen et al., 2016; Roscoe et al., 2014). The test is designed to evaluate general academic knowledge and includes a mix of science items, history, and literature items. Prior work has shown good reliability ( $\alpha$ 's = 0.72 - 0.81; Allen et al., 2016).

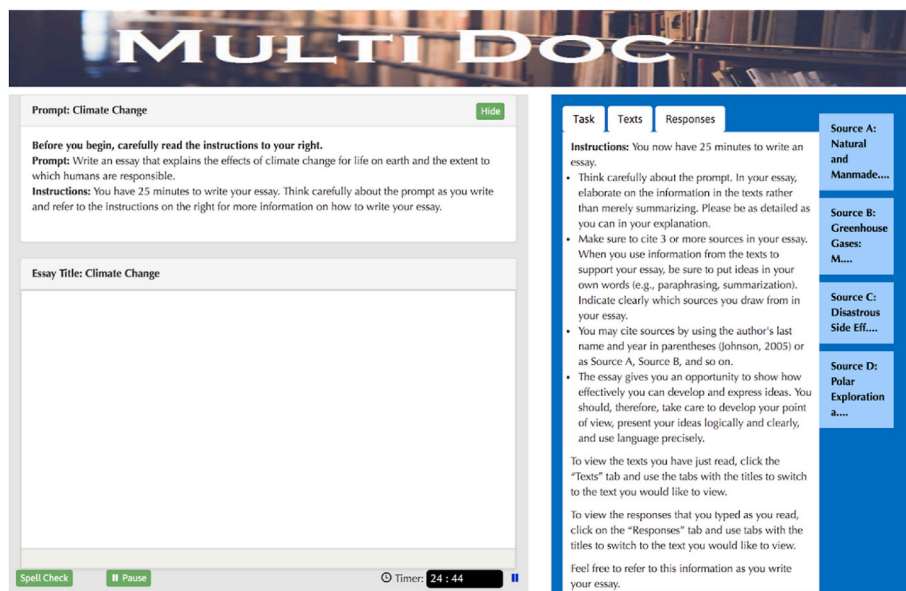


Fig. 1. iSTART's Multi-doc Integrated Essay Interface.

**2.3.4.4. Metacognitive strategies.** The Metacognitive Awareness of Reading Strategies Inventory (MARS; Mohktari & Reichard, 2002) is a common measure of knowledge of reading strategies. The MARS is a 30-item Likert survey. In this study, we used the modified MARS presented in Karimi (2015) that was adapted for use within a multiple-document context. In this 27-item version of the assessments, participants were asked the extent to which they used the strategy in the task. For example, an original item from the MARS is “I summarize what I read to reflect on important information in the text.”. The MARS-MD (Karimi, 2015) version was modified to “While reading these multiple texts, I summarized what I read to reflect on important information in them.” MARS-MD score was calculated by taking the average rating (1–5) on all 27 items.

## 2.4. Scoring

The two types of essays (independent essays and source-based integrated essays) were scored on different rubrics that have been developed and refined by writing experts. Both rubrics include a holistic score and multiple subscores. The rubrics and scales were developed and validated in the context of other research examining source-based writing and independent writing as separate tasks. Rather than trying to make the scores parallel, we elected to maintain the intended construct validity. However, prior work has shown some convergent validity as well as divergent validity across the two types of rubrics (e.g., Kim et al., 2021; Kyle & Crossley, 2016).

Although the integrated and independent essays were scored using different task-appropriate rubrics, the method of scoring and achieving reliability was the same. The expert raters were four graduate students with multiple years of experience as writing instructors. Rater pairs were trained to a high level of reliability on all metrics (i.e., all kappas >.80). Essays were randomly assigned to the pairs and each essay was then scored by two raters. If the two raters scored differed by more than 2, the scores were adjudicated by a third party. The final scores for each essay reflect the average score of the two raters.

### 2.4.1. Independent essays

Independent essays were assigned both a holistic score (1–6) as well as subscores (see Appendix D). Independent essay subscores include: (1) *introduction paragraph quality*, (2) *body paragraph quality*, (3) *conclusion paragraph quality*, (4) *organization*, (5) *topic and global cohesion*, (6) *grammar, style, and mechanics* (7) *voice*, (8) *word choice*, and (9) *sentence structure*. The holistic rubric is based on previous iterations of the SAT rubric (CollegeBoard) and has been used in several other studies of writing (e.g., Authors, 2015). Across raters, the reliability of ratings for scores (ICC) ranged from 0.71 to 0.86.

The average subscores (Tables 1 and 2) for the independent essays were moderately to strongly correlated ( $r_s = 0.42\text{--}0.84$ ) and there were no significant differences in the subscores across the two prompts (all  $p_s > .05$ ). There was a strong correlation between holistic score on the two independent essays ( $r = 0.63$ ). Thus, we used the combined holistic score (Images + Competition) to reflect a measure of general writing ability. This holistic score was strongly correlated with the summed subscores ( $r = 0.96$ ).

### 2.4.2. Integrated essays

Essays were assigned a holistic score (1–6) as well as subscores from 1 to 4 (see Appendix C). Integrated essay subscores include (1) *argumentation*, (2) *source use*, (3) *language sophistication*, and (4) *organization*. This rubric was developed based on existing writing rubrics (e.g., TOEFL) and consistent with the extant body of work in integrated writing (e.g., Martinez et al., 2015; van Ockenburg et al., 2019; Van dermeulen et al., 2020). Across raters, the reliability of ratings for scores (ICC) ranged from 0.65 to 0.82. Sample essays with explanations for their scores appear in Appendix E.

The average subscores scores and correlations for the two integrated

essays are shown in Tables 3 and 4. T-tests indicated no differences as a function of text set (all  $p_s > .05$ ). Consistent with the independent essays, the correlation between a combined holistic score and the summed subscores was strongly correlated ( $r = 0.96$ ).

## 2.5. Analytic approach

The unique and combined effects of individual differences were examined on integrated essay scores. Due to the varying ranges of scores (Table 5), z-scores were computed for each of the individual difference measures. We conducted a series of linear mixed effects models using the *lme4* package in R (Bates et al., 2014). The baseline model (m0) included participant nested within grade as a random factor and text set as a fixed factor. The additional models added the following fixed factors: constructed response prompt condition, general knowledge, Gates-MacGinitie Reading, Gates-MacGinitie Vocabulary, MARS-MD score, and independent essay score. Although preliminary analyses indicated no significant differences in integrated essay subscores across the two multiple-document text sets (global warming, cell phones), text set was included as a fixed effect in a baseline model (m0). To control for differences across the two grade levels (high school, college), we included participant nested within grade in the baseline model as well. Model 1 examined the effect of prompt condition (think-aloud, self-explanation, source evaluation). Each individual difference was then added in order from distal to proximal as the skill was assumed to related to the integrated writing task. Thus, general knowledge score was entered into the second model (m2) as it is the most general measure relative to the multiple-document inquiry task. In Model 3, we added Gates comprehension score, Gates vocabulary score, and metacognitive strategies (MARS-MD). These measures were assumed to reflect skills and strategies more immediately related to reading comprehension. It was predicted that those with better reading skills would be able to construct a more coherent and integrated mental model during the reading task, which would afford a higher quality essay. However, it was also assumed that a coherent mental model of the text topic would be necessary, but not sufficient for writing a high-quality integrated essay because integrated writing also involves separate writing skills. Thus, writing skill as measured by the independent essay scores was entered in Model 4. By adding the individual difference in this order, we were able to examine the extent to which writing skill positively contributes to integrated writing performance above and beyond the contribution of reading skill. Finally, we tested three two-way interactions across each of the three types of individual differences: general knowledge by writing skill, general knowledge by reading skill, and reading skill by writing skill. For simplicity, Gates comprehension score was used to reflect variability in reading skill. In the first set of analyses, holistic score for each integrated essay was used as the dependent variable<sup>1</sup>. We then conducted parallel analyses with each of the subscores.

## 3. Results

### 3.1. Preliminary analyses

Preliminary analyses showed good reliability for the individual difference measures (general knowledge:  $\alpha = 0.71$ ; GMRT:  $\alpha = 0.90$ ; GMVT:  $\alpha = 0.84$ ; MARS-MD:  $\alpha = 0.86$ ). These values were consistent with values from previous literature. Bivariate correlations (Table 5) reveal moderate correlations between the reading skills tests (Gates-MacGinitie Reading Comprehension and Gates-MacGinitie Vocabulary) and the independent essay scores, suggesting that these literacy skills are related, but at least somewhat independent. The correlations between both knowledge and the reading skill measures were stronger for the

<sup>1</sup> We also ran parallel analyses with summed subscores rather than holistic scores and found similar results.

**Table 1**  
Means, standard deviations, and correlations across subscores for images independent essays.

	M	SD	1	2	3	4	5	6	7	8	9
1. Images Introduction	3.65	0.85									
2. Images Body	3.66	0.85	.73**								
3. Images Conclusion	3.13	1.09	.56**	.71**							
4. Images Organization	3.46	0.95	.70**	.81**	.74**						
5. Images Cohesion	3.89	0.9	.72**	.79**	.66**	.77**					
6. Images Grammar	3.61	0.7	.66**	.58**	.44**	.64**	.63**				
7. Images Voice	3.57	0.73	.57**	.59**	.47**	.64**	.61**	.55**			
8. Images Word Choice	3.76	0.69	.70**	.61**	.42**	.62**	.63**	.75**	.63**		
9. Images Sentence Structure	3.74	0.68	.71**	.69**	.55**	.68**	.67**	.76**	.58**	.72**	
10. Images Holistic	3.46	0.99	.78**	.84**	.77**	.82**	.85**	.69**	.65**	.70**	.76**

Note. \*\**p* < .01.

**Table 2**  
Means, standard deviations, and correlations across subscores for competition independent essays.

	M	SD	1	2	3	4	5	6	7	8	9
1. Competition Introduction	3.69	0.86									
2. Competition Body	3.59	0.97	.73**								
3. Competition Conclusion	3.01	1.08	.58**	.74**							
4. Competition Organization	3.45	0.98	.73**	.87**	.78**						
5. Competition Cohesion	3.99	0.83	.77**	.73**	.63**	.81**					
6. Competition Grammar	3.67	0.85	.74**	.61**	.53**	.72**	.76**				
7. Competition Voice	3.54	0.84	.68**	.73**	.56**	.73**	.72**	.70**			
8. Competition Word Choice	3.69	0.76	.74**	.67**	.59**	.74**	.81**	.85**	.75**		
9. Competition Sentence Structure	3.71	0.81	.73**	.68**	.54**	.77**	.80**	.88**	.74**	.89**	
10. Competition Holistic	3.46	1.02	.78**	.86**	.83**	.88**	.80**	.77**	.74**	.78**	.80**

Note. \*\**p* < .01.

**Table 3**  
Means, standard deviations, correlations across subscores for global warming integrated essays.

	M	SD	1	2	3	4
1. GW Argumentation	2.70	0.70				
2. GW Source Use	2.59	0.78	.62**			
3. GW Language Sophistication	2.83	0.51	.53**	.44**		
4. GW Organization	2.63	0.69	.69**	.60**	.51**	
5. GW Holistic	2.82	0.95	.80**	.78**	.61**	.80**

Note. \*\**p* < .01.

**Table 4**  
Means, standard deviations, correlations across subscores for cell phones integrated essays.

	M	SD	1	2	3	4
1. CP Argumentation	2.62	0.76				
2. CP Source Use	2.53	0.74	.70**			
3. CP Language Sophistication	2.75	0.61	.64**	.67**		
4. CP Organization	2.58	0.76	.72**	.74**	.68**	
5. CP Holistic	2.79	1.06	.80**	.86**	.80**	.86**

Note. \*\**p* < .01.

integrated essays than the independent essays. There were also

**Table 5**  
Correlations across individual differences measures and integrated essay score.

	Possible Range	M	SD	1	2	3	4	5
1. General Knowledge Test	0–30	16.66	4.51					
2. Gates Reading Test	0–48	26.46	9.08	.53**				
3. Gates Vocabulary Test	0–45	28.56	6.75	.69**	.69**			
4. MARSIMD	1–5	3.45	0.55	–0.04	0.08	–0.03		
5. Independent Essay Score	0–12	6.92	1.81	.38**	.34**	.39**	–0.08	
6. Integrated Essay Score	0–12	5.61	1.78	.55**	.48**	.52**	0.08	.64**

Note. \*\**p* < .01.

moderate to strong correlations between general knowledge and the two Gates tasks. Integrated essay performance was moderately to strongly correlated with general writing skill, reading skill, and general knowledge. Interestingly, the participants' self-reported strategy use (MARSIMD) was not correlated with any of the individual difference measures nor the integrated essay scores.

Although our focus was on high school students, we had expanded our recruitment efforts to college freshmen. Despite recruiting college freshmen who had just graduated high school in the previous spring, there were significant differences in performance across the college and high school students. *T*-tests revealed that college students performed significantly better on the general knowledge test, Gates-MacGinitie Reading, Gates-MacGinitie Vocabulary, and integrated essay score (Table 6). Although differences in MARSIMD score and independent essay score were in this same direction, these tests failed to reach significance. These differences are unsurprising if one assumes that the selectivity of college admissions is likely to ensure that the college sample reflects the higher performing high school students. As we did not have any a priori predictions related to grade and that our constructed response prompt was randomly assigned across grade level, we do not explore this further. However, participant nested within grade is included in subsequent analyses to account for this variance.

Preliminary analyses (ANOVAs) revealed that random assignment to constructed response prompt condition (think-aloud, self-explanation, source evaluation) yielded even groups in terms of participants'

**Table 6**  
Means, Standard Deviations, and *t*-test Results of Individual Difference Measures and Essay Scores as a Function of Grade Band.

	High School (n = 49)	College (n = 47)	<i>t</i>	<i>p</i>
	M (SD)	M (SD)		
General Knowledge Test (0–30)	15.10 (4.40)	18.36 (4.02)	-3.75*	0.00
Gates Reading Test (0–48)	23.45 (8.44)	29.73 (8.68)	-3.55*	0.00
Gates Vocabulary Test (0–45)	26.69 (6.97)	30.60 (5.93)	-2.94*	0.00
MARSI-MD (1–5)	3.38 (0.51)	3.52 (0.59)	-1.18	0.24
Independent Essay Score (0–12)	6.66 (1.75)	7.20 (1.86)	-1.43	0.15
Integrated Essay Score (0–12)	5.11 (1.51)	6.16 (1.91)	-2.92*	0.00

general knowledge, reading skill, reading strategies, and independent essay score (all *F*s < 1.00).

3.2. Predicting integrated writing scores

3.2.1. Holistic score

Model summaries appear in the top rows of Table 7. Likelihood ratio tests indicate differences between each model and a reduced model. Significant chi-square ( $\chi^2$ ) tests indicate that adding the additional variable(s) improved fit as compared to the previous model. Coefficients for each model are presented in Table 7.

The addition of constructed response condition (think-aloud, self-explanation, or source evaluation) to Model 1 did not improve fit beyond the grade, participant, and text set factors included in the baseline model. Analysis of Model 2 showed that adding general knowledge to the model improved model fit and indicated that general knowledge was a strong, positive predictor of essay scores. Model 3 included the measures related to reading: Gates Reading, Gates Vocabulary, and MARSI-MD. Although beta weights indicated a positive relation between all three reading scores, the reading measures did not significantly improve model fit compared to Model 2. To test the contribution of general writing ability to integrated essay score, independent essay scores were added in Model 4. This model improved model fit as compared to the model that included text set, general knowledge, Gates Reading, Gates Vocabulary, and the MARSI-MD. Coefficients revealed that general knowledge remained a significant predictor ( $\beta = 0.23$ ), but that independent essay score was a stronger predictor of integrated essay score ( $\beta = 0.43$ ), even when controlling for the contributions of reading skills and self-reported strategy use.

Finally, three interactions were tested in comparison to Model 4: general knowledge by writing skill (m5a), general knowledge by reading skill (m5b), and reading skill by writing skill (m5c). None of these interactions improved model fit relative to the main effect model, m4 (m5a:  $\chi^2 < 0.03$ ,  $p = .87$ ; m5b:  $\chi^2 = 1.67$ ,  $p = .20$ ; m5c:  $\chi^2 = 0.30$ ,  $p = .58$ ).

3.2.2. Subscores

Holistic score was strongly correlated with each of the four subscores (Tables 3 and 4). However, holistic score may tap into overall writing quality rather than specific features of the essays that reflect content comprehension and integration. Thus, we conducted analyses to evaluate the extent to which the individual differences predicted the four subscores (argumentation, source use and integration, language sophistication, and organization). These analyses more closely reflect what has been done in previous work related to multiple document comprehension and source-based writing.

For all four subscores, adding general knowledge to the model (m2) improved model fit (Tables 8–11). Adding the reading skill measures

**Table 7**  
Likelihood ratio tests and coefficients for mixed-effects models predicting integrated essay scores.

	Model 1			Model 2			Model 3			Model 4		
	AIC	$\chi^2$	<i>p</i>	AIC	$\chi^2$	<i>p</i>	AIC	$\chi^2$	<i>p</i>	AIC	$\chi^2$	<i>p</i>
Intercept	506.38	529.04	1.04	477.94	503.83	30.44	475.22	510.82	8.72	443.52	482.35	33.70
Text Set	2.92	0.30	9.84	2.89	0.15	19.06	2.86	0.14	20.86	2.83	0.12	23.92
Condition (self-explain)	0.03	0.10	0.28	0.03	0.10	0.28	0.03	0.10	0.28	0.03	0.10	0.28
Condition (source)	0.13	0.22	0.59	0.15	0.19	0.78	0.11	0.19	0.56	0.01	0.16	0.06
General Knowledge	-0.09	0.22	-0.42	0.01	0.19	0.03	0.00	0.19	0.02	-0.09	0.16	-0.59
Gates Reading				0.48	0.08	6.00	0.32	0.11	3.05	0.23	0.09	2.54
Gates Vocabulary							0.16	0.11	1.49	0.11	0.09	1.23
MARSI-MD							0.13	0.13	1.01	0.06	0.11	0.54
Independent Essay Score							0.07	0.08	0.97	0.43	0.07	1.60
												0.07
												6.09
												< 0.01**

Note. \*\**p* < .01.



(m3) improved model fit for the argumentation score (Table 8) and source use score (Table 9) but did not improve model fit for the language sophistication score (Table 10) nor the organization score (Table 11). This finding supports the notion that argumentation and source are related to *comprehension* processes, whereas language use and organization of ideas may be more related to production and composition. Critically, consistent with what was found for the holistic score, these effects were subsumed by independent essay score representing writing skills. That is, once independent essay scores was added to the models, it became the strongest predictor of subscores – even for those measures assumed to reflect comprehension. In these best fit models, independent essay score was the only significant predictor for the source use, language sophistication, and organization scores.

4. Discussion

Multiple-document integrated reading and writing tasks are a common practice in everyday life and in the classroom. While there has been an abundance of studies examining multiple-document comprehension through integrated writing, this work has been criticized because rather than treating the task as a hybrid task (e.g., Spivey, 1989), as it is in the study of synthesis writing, multiple-document comprehension researchers tend to view the essay as a means-to-an-end (making inferences about the processes that occurred during reading) without considering how students’ writing ability might impact the essays (e.g., Du & List, 2020; Primor & Katzir, 2018; Wiley et al., 2018). This study highlights this issue by showing that students’ general writing ability strongly impacts the quality of a multiple-document source-based integrated essay above and beyond the contributions of strategies and skills related to reading comprehension. In addition to evaluating the contribution of individual differences in reading and writing on overall writing quality (i.e., holistic score), we also found that general writing ability was a strong predictor of subscores that have been used as proxies for comprehension in prior studies of multiple documents comprehension (e.g., argumentation, source use, and organization). This further highlights the need to measure and account for writing skill when making assumptions about comprehension processes in multiple document studies.

We used a constructed response prompt manipulation to encourage deeper comprehension through two strategies (i.e., self-explanation and source evaluation) that have been shown to support integration and comprehension in both single and multiple-document scenarios. Neither of these strategy prompts yielded differences in the integrated essay quality relative to a think-aloud control. On the one hand, this finding is inconsistent with the existing work showing benefits of self-explanation and source evaluation. On the other hand, it is consistent with the idea that single-dose prompt manipulations may be insufficient for improving strategy use (e.g., McNamara et al., 2004, 2017). Strategy interventions, such as iSTART (McNamara et al., 2004) for self-explanation and Sourcer’s Apprentice (Britt & Aglinskas, 2002) and SEEK (Sanchez et al., 2006) for source evaluation include multi-session interventions that include explicit instruction, training, and feedback. Thus, these findings encourage two further lines of the work; the first is to further investigate how the prompt manipulation may have impacted the nature of the constructed responses and the extent to which this might mediate differences in essay content or quality. The second is to provide more in-depth strategy instruction for both self-explanation and source evaluation to enhance the ability to examine the impact of these strategies on both comprehension and production.

In terms of the individual differences, correlational analyses revealed that general knowledge, reading skill, vocabulary, and general writing ability were positively related to the quality of integrated essays. Interestingly, MARSJ-MD was not related to the other individual difference measures. There are a few possible explanations. First, there is a potential issue of using self-report. Some research has shown that self-reported measures of strategy use are only weakly related to objective

Table 8 Likelihood ratio tests and coefficients for mixed-effects models predicting argumentation subscores.

	Model 1			Model 2			Model 3			Model 4		
	AIC	BIC	$\chi^2$	p	AIC	BIC	$\chi^2$	p	AIC	BIC	$\chi^2$	p
Intercept	2.69	0.17	15.88	0.00	2.66	0.11	24.42	0.00	2.64	0.11	24.65	<0.01
Text Set	0.07	0.08	0.96	0.34	0.07	0.08	0.96	0.34	0.07	0.08	0.96	0.34
Condition (self-explain)	0.02	0.16	0.16	0.87	0.04	0.14	0.28	0.78	0.06	0.14	0.41	0.68
Condition (source)	-0.21	0.16	-1.37	0.18	-0.15	0.14	-1.08	0.28	-0.13	0.14	-0.90	0.37
General Knowledge					<b>0.29</b>	<b>0.06</b>	<b>5.00</b>	< <b>0.01**</b>	<b>0.17</b>	<b>0.08</b>	<b>2.21</b>	<b>0.03*</b>
Gates Reading					0.03	0.08	0.33	0.75	0.03	0.08	0.33	0.75
Gates Vocabulary					0.16	0.09	1.70	0.09	0.16	0.09	1.70	0.18
MARSJ-MD					0.11	0.06	1.97	0.05	0.11	0.06	1.97	0.05
Independent Essay Score									<b>0.29</b>	<b>0.05</b>	<b>5.27</b>	< <b>0.01**</b>

Note. \*p < .05, \*\*p < .01.

**Table 9**  
Likelihood ratio tests and coefficients for mixed-effects models predicting source use subscores.

	Model 1			Model 2			Model 3			Model 4		
	AIC	BIC	$\chi^2$	AIC	BIC	$\chi^2$	AIC	BIC	$\chi^2$	AIC	BIC	$\chi^2$
<i>B</i>	400.36	423.01	2.89	380.30	406.19	22.05	376.77	412.37	9.53	352.46	391.30	26.31
Intercept	2.50	0.24	10.49	2.47	0.15	16.73	2.47	0.12	19.88	2.51	0.12	21.09
Text Set	0.06	0.07	0.82	0.06	0.07	0.82	0.06	0.07	0.82	0.06	0.07	0.82
Condition (self-explain)	0.17	0.16	1.08	0.19	0.15	1.26	0.17	0.15	1.13	0.11	0.14	0.80
Condition (source)	-0.06	0.16	-0.38	0.00	0.15	-0.02	0.01	0.15	0.05	-0.06	0.14	-0.41
General Knowledge				<b>0.29</b>	<b>0.06</b>	<b>4.53</b>	<b>0.18</b>	<b>0.08</b>	<b>2.13</b>	0.12	0.08	1.52
Gates Reading							0.09	0.09	1.00	0.05	0.08	0.67
Gates Vocabulary							0.12	0.10	1.25	0.08	0.09	0.89
MARSI-MD							0.08	0.06	1.24	0.09	0.06	1.69
Independent Essay Score										<b>0.28</b>	<b>0.06</b>	<b>4.59</b>
												<b>&lt; 0.01**</b>

Note. \**p* < .05, \*\**p* < .01.

**Table 10**  
Likelihood ratio tests and coefficients for mixed-effects models predicting language use subscores.

	Model 1			Model 2			Model 3			Model 4		
	AIC	BIC	$\chi^2$	AIC	BIC	$\chi^2$	AIC	BIC	$\chi^2$	AIC	BIC	$\chi^2$
<i>B</i>	400.36	423.01	2.89	380.30	406.19	22.05	376.77	412.37	9.53	352.46	391.30	26.31
Intercept	2.70	0.19	13.88	2.69	0.15	17.83	2.70	0.14	19.86	2.72	0.13	20.27
Text Set	0.08	0.06	1.25	0.08	0.06	1.25	0.08	0.06	1.25	0.08	0.06	1.25
Condition (self-explain)	0.14	0.11	1.30	0.15	0.11	1.41	0.12	0.11	1.14	0.09	0.10	0.86
Condition (source)	0.02	0.11	0.22	0.05	0.11	0.49	0.04	0.11	0.40	0.01	0.10	0.06
General Knowledge				<b>0.13</b>	<b>0.05</b>	<b>2.91</b>	0.09	0.06	1.52	0.06	0.06	0.98
Gates Reading							0.08	0.06	1.27	0.06	0.06	1.02
Gates Vocabulary							0.01	0.07	0.12	0.06	0.07	-0.26
MARSI-MD							0.02	0.04	0.38	0.03	0.04	0.67
Independent Essay Score										<b>0.17</b>	<b>0.05</b>	<b>3.63</b>
												<b>&lt; 0.01**</b>

Note. \*\**p* < .01.

**Table 11**  
Likelihood ratio tests and coefficients for mixed-effects models predicting organization subscores.

	Model 1			Model 2			Model 3			Model 4			
	AIC	BIC	$\chi^2$	AIC	BIC	$\chi^2$	AIC	BIC	$\chi^2$	AIC	BIC	$\chi^2$	
Intercept	2.70	0.19	13.88	2.69	0.15	17.83	2.70	0.14	19.86	2.72	0.13	20.27	
Text Set	0.08	0.06	1.25	0.08	0.06	1.25	0.08	0.06	1.25	0.08	0.06	1.25	
Condition (self-explain)	0.14	0.11	1.30	0.15	0.11	1.41	0.12	0.11	1.14	0.09	0.10	0.86	
Condition (source)	0.02	0.11	0.22	0.05	0.11	0.49	0.04	0.11	0.40	0.01	0.10	0.06	
General Knowledge				<b>0.13</b>	<b>0.05</b>	<b>2.91</b>	0.09	0.06	1.52	0.06	0.06	0.98	
Gates Reading							0.08	0.06	1.27	0.21	0.06	1.02	
Gates Vocabulary							0.01	0.07	0.12	-0.02	0.07	-0.26	
MARSI-MD							0.02	0.04	0.38	0.03	0.04	0.67	
Independent Essay Score										<b>0.17</b>	<b>0.05</b>	<b>3.63</b>	
													<b>&lt; 0.01**</b>

Note. \*\*p < .01.

literacy measures (e.g., Cromley & Azevedo, 2006). However, others have argued that self-report measures may be valid if they are grounded to a specific task and judgments are made retrospectively (Bråten et al., 2020; McNamara, 2011). Whereas the original MARSI is designed to evaluate students' general awareness of metacognitive strategies, MARSI-MD is a grounded assessment, in which students are asked to reflect on what they did when reading these texts during the targeted reading task. As there were two MD tasks, each conducted on separate days (i.e., Sessions 1 and 2), participants completed the MARSI-MD in Session 3 of the study, rather than immediately after completing the MD tasks. This design was decidedly optimal because administering the MARSI-MD immediately after Session 1 may have biased performance in Session 2 and administering immediately after Session 2 may have led to responses that were specific to only Session 2. Nonetheless, administering MARSI-MD on day three potentially reduced its validity as a grounded measure if participants did not remember which strategies that they had used during sessions one and two. Additionally, participants may not have perceived the strategies specified in the MARSI-MD as relevant to the particular reading and writing tasks engaged in this study. Indeed, the low level of variability in responses suggests this may be the case. Finally, the original MARSI yields three subscores that have not been validated in the MARSI-MD. A single MARSI-MD score may obscure more subtle differences and in turn reduce its predictive validity. Notably, however, participants' responses on the MARSI-MD were weakly predictive of argumentation subscores ( $r = 0.19$ ), somewhat suggestive of relations between comprehension strategies and the ability to develop coherent arguments in integrated essays. However, given the multiple concerns regarding the MARSI-MD, as well as the predominance of null correlations, this singular, modest finding should be interpreted with caution. Future research with larger samples will be necessary to further explore the role of metacognitive strategies through additional or alternative methods of capturing strategic processing.

We also found that reading skill (as measured by the Gates Reading and Gates Vocabulary) did not contribute significantly to the best fitting models explaining the quality of the integrated writing essays for both overall holistic score as well as for the four subscores. This would be unexpected given that reading skill is an established predictor in multiple-document research (Barzilai & Strømsø, 2018). One possible explanation is that the measure of vocabulary knowledge (i.e., the Gates vocabulary score) was strongly correlated with both comprehension and general knowledge ( $r_s = .69$ ). Vocabulary knowledge is both a facilitator and outcome – a larger vocabulary facilitates reading which, in turn, supports knowledge gain, so it is unsurprising that those with a greater breadth of vocabulary also demonstrate greater general academic knowledge (McCarthy & McNamara, 2021; Wang et al., 2021). Thus, the contributions of variability in both Gates scores may have failed to significantly account for variance because of an inherent overlap in general knowledge and reading skill.

The results suggest that general proficiencies in text production and conveying one's ideas in writing are critical contributors to the quality of integrated writing, above and beyond the contribution of abilities related to comprehension. From an applied standpoint, this is important because most of the interventions for multiple-document inquiry focus on the comprehension aspects of the task, and particularly on evaluating the credibility of the source. Although these processes are undoubtedly important, such interventions overlook the need to support students' ability to translate their ideas onto the page and organize them coherently. From a more theoretical perspective, this study suggests the need to better understand the relations between comprehension and production and highlights a need for the development of an integrated model of reading and writing (e.g., McNamara and Allen, 2017; McNamara, Roscoe, Allen, Balyan, & McCarthy, 2019).

Researchers have identified a number of complex processes involved in successful multiple-document reading and writing (e.g., Barzilai & Strømsø, 2018; Britt et al., 2017; List & Alexander, 2019; Rouet et al., 2017). However, much of the research has focused on the

comprehension processes that occur during reading. The results of the present study clearly suggest that more emphasis must be placed on understanding the writing skills and processes involved as learners' attempt to use their mental model of the document content and translate it to the page.

#### 4.1. Limitations & future directions

The present study provides new insights into the relations between reading and writing in an integrated writing task. However, there are some limitations in this work that should be considered and drive future research. The first is related to the outcome measure. While some of the subscores are related to comprehension of the content (e.g., source use), others draw more heavily upon other aspects of writing and composition. For example, argumentation score reflects not only comprehension of the materials, but also discipline-specific argumentation skills (e.g., Goldman et al., 2016). Both language and organization subscores are likely to reflect more general composition skill rather than understanding of the text content. It is not our intent to argue that existing multiple-document comprehension rubrics do not include these dimensions, but they are not often made explicit. This was an intentional decision to highlight the fact that integrated essays are a writing task and not a pure window into comprehension as is often assumed in research on multiple-document comprehension. A writing focused rubric is also more consistent with the type of evaluation that a student might receive as part of a multiple-document activity assigned in a classroom.

An additional limitation in this study is our wording in the verbal protocol instructions, which indicated that the participant would be reading with the goal of "answering comprehension questions." Our intent was to provide a relatively ambiguous reading goal, rather than providing the essay instruction prior to reading. Given the robust body of work on task relevance in single and multiple document comprehension research, it is likely that the effects in the present study would be moderated by different task instructions. Future work should consider this. Notably, we did not include comprehension questions in the study. Many multiple-document comprehension studies include comprehension questions such as *inference verification tests* (e.g., Bråten & Strømsø, 2011). These tests allow researchers to probe for evidence of specific intra- and inter-textual inferences that may not emerge in written products. We did not include these types of items in this study because including both essay and comprehension questions poses challenges for interpreting results. Providing comprehension questions prior to writing can cue participants to concepts and connections they may not have made on their own. Conversely, a writing task is likely to drive additional reflection and integration that may bias performance on subsequent comprehension questions. Thus, more systematic work should be conducted to identify the specific skills and knowledge that are being measured by these various tasks so that researchers can better identify when to use these various measures in their own research.

Second, this study focused on domain-general individual difference measures. That is, the tests were designed to evaluate general academic knowledge, reading skill, and writing ability. However, multiple-document integrated writing tasks are likely to be dependent not only on these general factors, but also on knowledge, skills, and strategies that are specific to the topic and type of task at hand (McCarthy & McNamara, 2021). For example, general academic knowledge is superseded by relevant topic knowledge when learners are engaging in complex, multiple-document comprehension tasks (e.g., Wang et al., 2021). Topic knowledge is a known factor in both reading comprehension (e.g., Alexander et al., 1994; Ozuru et al., 2009) and writing quality (e.g., Hammann & Stevens, 2003; Langer, 1984). In the current study, we did not collect measures of topic-specific knowledge for each of the four essay topics. It is likely that the variation seen in essay quality across the essay type and essay topics can be at least partially explained by topic general knowledge. Similarly, general reading skill as measured

by reading and answering isolated passages may only be one part of deeper meaning-making and integration in real-world learning tasks (e.g., Magliano et al., 2018; Sabatini et al., 2019) and the specific skills and strategies at play may vary depending on the type of text or the context in which the text is being read (e.g., McCarthy, 2020; Rouet et al., 2017; van den Broek et al., 2001). In a similar fashion, the independent essay score served as a proxy for general writing ability. Writing is comprised of a variety of lower and higher order skills and strategies. This study suggests a need to more carefully examine how these component skills might predict performance on integrated writing.

In the current study, students were prompted to write an explanation of the socioscientific issues raised in the text sets. Prior work in multiple-documents reading and writing has shown that different essay prompts (e.g., write to inform vs. write to argue) may affect the way that the learner interprets that task (e.g., Britt et al., 2017; List et al., 2019) and may change the nature of the processes and strategies that emerge during reading and writing and the quality of the final written products (Gil et al., 2010; Vandermeulen et al., 2020; Wiley & Voss, 1999). In the same vein, the ability to translate and organize one's ideas into writing is an important foundation, but students also need to engage higher-order disciplinary argumentation skills that may be specific to the goals and expectations of the discipline or genre in question (e.g., Goldman, 2018; Goldman et al., 2016). Our results suggest that future work in multiple-document integrated reading and writing should examine the contributions and potential interactions between both general proficiencies and more text, task, and discipline-specific knowledge and skills.

Finally, there are a number of other individual differences beyond reading and writing skills that influence the way that text is processed and how integrated tasks might unfold. For example, several studies have demonstrated that attitudes, beliefs, and epistemologies can predict the extent to which readers seek out, use, and integrate information, especially in the context of controversial topics (e.g., Bråten & Strømsø, 2020, 2010; Kunda, 1990; Maier & Richter, 2016; McCrudden & Sparks, 2014; Wiley, Griffin, Steffens, & Britt, 2020). Future work should consider how the cold cognitive processes described in this paper interplay with warm, affective individual differences related to motivation and engagement (List & Alexander, 2019).

#### 4.2. Conclusions

Multiple-document inquiry tasks in the classroom are an attempt to emulate the types of problem solving and learning that students must engage in to deeply understand and interact with the world around them. These tasks require learners to evaluate and integrate information from a variety of sources in order to develop and convey complex ideas and demand a variety of knowledge, skills, and strategies. The past few decades have elucidated the processes and factors related to multiple-document comprehension. The next few decades should be dedicated to extending our understanding of how these processes come together as students read and write about complex topics. A thorough investigation of this area would include the processes and factors involved in integrated writing quality as well as the interaction between comprehension measures and writing outcome measures.

#### Author statement

Kathryn S. McCarthy: Conceptualization; Methodology; Analysis; Writing; Supervision. Eleanor F. Yan: Investigation; Analysis; Writing. Laura K. Allen: Conceptualization; Methodology; Writing; Supervision. Alison N. Sonia: Analysis; Writing. Joseph P. Magliano: Conceptualization; Methodology; Writing; Supervision. Danielle S. McNamara: Conceptualization; Methodology; Analysis; Writing; Supervision.



**Declarations**

The authors declare no conflicts of interest.

R305A180144 and R305A190063) as well as the Office of Naval Research (Grants: N00014-20-1-2627, N00014-19-1-2424, and N00014-20-2623). The views expressed are those of the authors and do not necessarily reflect the views of the Department of Education, IES, or the Office of Naval Research.

**Acknowledgements**

This research was made possible in part by grants from IES (Grants

**Appendix A. Integrated Essay Instructions**

*Essay Instructions (above essay entry box, see Fig. 1)*

You now have 25 min to write an essay.

**Prompt:** Write an essay that explains the effects of climate change for life on earth and the extent to which humans are responsible.

**Instructions:** Think carefully about the prompt. In your essay, elaborate on the information in the texts rather than merely summarizing. Please be as detailed as you can in your explanation.

*Instructions (right side of screen, see Fig. 1)*

Instructions: You now have 25 min to write an essay.

- Think carefully about the prompt. In your essay, elaborate on the information in the texts rather than merely summarizing. Please be as detailed as you can in your explanation.
- Make sure to cite 3 or more sources in your essay. When you use information from the texts to support your essay, be sure to put ideas in your own words (e.g., paraphrasing, summarization). Indicate clearly which sources you draw from in your essay.
- You may cite sources by using the author’s last name and year in parentheses (Johnson, 2005) or as Source A, Source B, and so on.
- The essay gives you an opportunity to show how effectively you can develop and express ideas. You should, therefore, take care to develop your point of view, present your ideas logically and clearly, and use language precisely.

To view the texts you have just read, click the “texts” tab and use the tabs with the titles to switch to the text you would like to view.

To view the responses that you typed as you read, click on the “Responses” tab and use tabs with the titles to switch to the text you would like to view.

Feel free to refer to this information as you write your essay.

**Appendix B. Independent Essay Instructions**

You will now have 25 min to write an essay on the prompt below. The essay gives you an opportunity to show how effectively you can develop and express ideas. You should, therefore, take care to develop your point of view, present your ideas logically and clearly, and use language precisely. Think carefully about the issue presented in the following excerpt and the assignment below. Plan and write an essay in which you develop your point of view on this issue. Support your position with reasoning and examples taken from your reading, studies, experience, or observations.

(Images and Impressions).

All around us appearances are mistaken for reality. Clever advertisements create favorable impressions but say little or nothing about the products they promote. In stores, colorful packages are often better than their contents. In the media, how certain entertainers, politicians, and other public figures appear is sometimes considered more important than their abilities. All too often, what we think we see becomes far more important than what really is. Do images and impressions have a positive or negative effect on people?

(Competition or Cooperation).

While some people promote competition as the only way to achieve success, others emphasize the power of cooperation. Intense rivalry at work or play or engaging in competition involving ideas or skills may indeed drive people either to avoid failure or to achieve important victories. In a complex world, however, cooperation is much more likely to produce significant, lasting accomplishments. Do people achieve more success by cooperation or by competition?

**Appendix C. Integrated Essay Rubric**

	1	2	3	4
<b>Argumentation</b>	<p><b>The essay:</b></p> <ul style="list-style-type: none"> <li>● does not discuss the sides of the argument nor provide a position</li> <li>● provides no claims to support a position</li> <li>● provides no evidence to support the side(s) or position</li> </ul>	<p><b>The essay:</b></p> <ul style="list-style-type: none"> <li>● discusses the side(s) of the argument but does not provide a position</li> <li>● discusses the side(s) by providing 1+ relevant or accurate claims</li> <li>● supports the side(s) by providing 1+ relevant or accurate pieces of evidence</li> </ul>	<p><b>The essay:</b></p> <ul style="list-style-type: none"> <li>● discusses the sides of the argument and implicitly states a position</li> <li>● supports the position by providing 1–2 relevant and accurate claims</li> <li>● supports the position and claims by providing 1–2 relevant and accurate pieces of evidence</li> </ul>	<p><b>The essay:</b></p> <ul style="list-style-type: none"> <li>● discusses the sides of the argument and explicitly states a position</li> <li>● supports the position by providing 3+ relevant and accurate claims</li> <li>● supports the position and claims by providing 3+</li> </ul>

(continued on next page)

(continued)

	1	2	3	4
<b>Source Use and Inferencing</b>	<p><b>The essay:</b></p> <ul style="list-style-type: none"> <li>● does not refer to any of the provided sources</li> <li>● does not synthesize information from within or across the sources provided</li> <li>● fails to reference concepts described in the source material</li> </ul>	<p><b>The essay:</b></p> <ul style="list-style-type: none"> <li>● implicitly refers to one or more of the outside sources</li> <li>● generates explicit or implicit connections from information within sources but fails to generate any connections across the multiple sources</li> <li>● relies heavily on direct quotes or paraphrases of the source material</li> </ul>	<p><b>The essay:</b></p> <ul style="list-style-type: none"> <li>● explicitly refers to one or more of the outside sources</li> <li>● generates explicit connections from information within sources but generates few explicit or implicit connections across the multiple sources</li> <li>● provides interpretations of sources that go somewhat beyond reiteration/paraphrase of the material</li> </ul>	<p>relevant and accurate pieces of evidence</p> <p><b>The essay:</b></p> <ul style="list-style-type: none"> <li>· explicitly refers to the majority of outside sources</li> <li>· synthesizes information from both within and across the referenced sources</li> <li>· provides deep interpretations of sources that go far beyond simple reiteration/paraphrase of the material</li> </ul>
<b>Language Sophistication</b>	<p><b>The essay:</b></p> <ul style="list-style-type: none"> <li>● demonstrates low lexical and syntactic sophistication and little to no variety</li> <li>● generally does not use appropriate word choice</li> <li>● contains a number of spelling, grammar, and mechanics errors that render portions of the text difficult to understand</li> </ul>	<p><b>The essay:</b></p> <ul style="list-style-type: none"> <li>● demonstrates little lexical and syntactic sophistication and little variety</li> <li>● contains some inappropriate word choices</li> <li>● shows some understanding of English spelling, grammar, and mechanics, but contains numerous errors</li> </ul>	<p><b>The essay:</b></p> <ul style="list-style-type: none"> <li>● demonstrates lexical and syntactic sophistication but little variety in word choice and syntax</li> <li>● generally uses appropriate word choices</li> <li>● shows an understanding of English spelling, grammar, and mechanics, but may contain some errors</li> </ul>	<p><b>The essay:</b></p> <ul style="list-style-type: none"> <li>● demonstrates lexical and syntactic sophistication and variety in word choice and syntax</li> <li>● uses consistently appropriate word choices</li> <li>● demonstrates a command of English spelling, grammar, and mechanics, containing few to no errors</li> </ul>
<b>Organization</b>	<p><b>The essay:</b></p> <ul style="list-style-type: none"> <li>● generally lacks a logical sequence of thought</li> <li>● lacks an appropriate organizational structure</li> <li>● is not coherent and lacks important cohesive elements</li> </ul>	<p><b>The essay:</b></p> <ul style="list-style-type: none"> <li>● sometimes deviates from logical structure and lacks introduction or concluding statements</li> <li>● contains some evidence of organization but lacks important transitions between central ideas and paragraphs</li> <li>● is somewhat coherent but lacks important cohesive elements</li> </ul>	<p><b>The essay:</b></p> <ul style="list-style-type: none"> <li>● follows a logical structure but lacks explicit introduction or concluding statements</li> <li>● contains evidence of organization but may lack some appropriate transitions between paragraphs</li> <li>● is coherent and generally makes appropriate use of cohesive devices to signal connections between ideas</li> </ul>	<p><b>The essay:</b></p> <ul style="list-style-type: none"> <li>● follows a logical structure, beginning with an introduction and ending with concluding statements</li> <li>● is well-organized and maintains sense of flow throughout the paragraphs</li> <li>● is coherent and makes appropriate use of cohesive devices to signal connections between ideas</li> </ul>

Holistic Essay Quality: \_\_\_\_/6.

- 1 Very Poor
- 2 Poor
- 3 Fair
- 4 Good
- 5 Very Good
- 6 Excellent

**Appendix D. Independent Essay Rubric**

*Independent Essay Rubric: Subscores*

Read each essay carefully and then assign a score on each of the points below. For the following evaluations, you will need to use a grading scale between 1 (minimum) and 6 (maximum).

We present here a description of the grade as a guide using the example of *does not meet the set criterion in any way* versus *meets the set criterion in every way*. For example, a grade of 1 would relate to not meeting the criterion in any way, and a grade of 4 would relate to somewhat meeting the criterion. The distance between each grade (e.g., 1–2, 3–4, 4–5) should be considered equal. Thus, a grade of 5 (*meets the criterion*) is as far above a grade of 4 (*somewhat meets the criterion*) as a grade of 2 (*does not meet the criterion*) is above a grade of 1 (*does not meet the criterion in any way*).

Score	Definition
1	Does not meet the criterion in any way
2	Does not meet the criterion
3	Almost meets the criterion but not quite
4	Meets the criterion but only just
5	Meets the criterion
6	

(continued on next page)

(continued)

Score	Definition
Meets the criterion in every way	
Subscore	Score (1–6)
1. Introduction	
1.1 <i>Introduction overall score</i>	
The writer demonstrates mastery in meeting the goals of an introduction (e.g., presenting a topic, providing a purpose, clearly stating a thesis, and previewing arguments)	
2. Body	
2.1 <i>Body overall score</i>	
The writer demonstrates mastery in meeting the goals of body arguments (e.g., transition between arguments, using topic sentences, supporting arguments with evidence, and maintaining a flow throughout the arguments)	
3. Conclusion	
3.1 <i>Conclusion overall score</i>	
The writer demonstrates mastery in meeting the goals of a conclusion (e.g., summarizing the essay, re-establishing the significance of discussion, capturing the reader's attention, and effectively closing the essay.)	
4. Text Qualities	
4.1 <i>Organization</i>	
The essay follows a logical structure, beginning with the introduction, through the arguments and evidence presented in the body arguments, and to the conclusion.	
4.2 <i>On-Topic/Global Cohesion</i>	
The details presented throughout the essay support the thesis and do not stray from the prompt and the main ideas and organizing principles presented in the introduction.	
4.3 <i>Grammar, syntax, and mechanics</i>	
The writer employs correct Standard American English, avoiding errors in grammar, syntax, and mechanics; the essay conveys strong control of the standard conventions of writing.	
5. Stylistic Considerations	
5.1 <i>Voice/Personality</i>	
The writer is expressive, engaging, and sincere, with a strong sense of audience.	
5.2 <i>Word Choice</i>	
The writer is precise and effective in word choice.	
5.3 <i>Sentence Structure</i>	
Sentence patterns are varied effectively, enhancing the quality of the essay.	

*Independent Essay Rubric: Holistic Score*

After reading each essay and completing the analytical rating form, assign a holistic score based on the rubric below. For the following evaluations you will need to use a grading scale between 1 (minimum) and 6 (maximum). As with the analytical rating form, the distance between each grade (e.g., 1–2, 3–4, 4–5) should be considered equal.

**SCORE OF 6:** An essay in this category demonstrates clear and consistent mastery, although it may have a few minor errors. A typical essay effectively and insightfully develops a point of view on the issue and demonstrates outstanding critical thinking, using clearly appropriate examples, reasons, and other evidence to support its position is well organized and clearly focused, demonstrating clear coherence and smooth progression of ideas exhibits skillful use of language, using a varied, accurate, and apt vocabulary demonstrates meaningful variety in sentence structure is free of most errors in grammar, usage, and mechanics.

**SCORE OF 5:** An essay in this category demonstrates reasonably consistent mastery, although it will have occasional errors or lapses in quality. A typical essay effectively develops a point of view on the issue and demonstrates strong critical thinking, generally using appropriate examples, reasons, and other evidence to support its position is well organized and focused, demonstrating coherence and progression of ideas exhibits facility in the use of language, using appropriate vocabulary demonstrates variety in sentence structure is generally free of most errors in grammar, usage, and mechanics.

**SCORE OF 4:** An essay in this category demonstrates adequate mastery, although it will have lapses in quality. A typical essay develops a point of view on the issue and demonstrates competent critical thinking, using adequate examples, reasons, and other evidence to support its position is generally organized and focused, demonstrating some coherence and progression of ideas exhibits adequate but inconsistent facility in the use of language, using generally appropriate vocabulary demonstrates some variety in sentence structure has some errors in grammar, usage, and mechanics.

**SCORE OF 3:** An essay in this category demonstrates developing mastery, and is marked by ONE OR MORE of the following weaknesses: develops a point of view on the issue, demonstrating some critical thinking, but may do so inconsistently or use inadequate examples, reasons, or other evidence

to support its position is limited in its organization or focus, or may demonstrate some lapses in coherence or progression of ideas displays developing facility in the use of language, but sometimes uses weak vocabulary or inappropriate word choice lacks variety or demonstrates problems in sentence structure contains an accumulation of errors in grammar, usage, and mechanics.

**SCORE OF 2:** An essay in this category demonstrates little mastery, and is flawed by ONE OR MORE of the following weaknesses: develops a point of view on the issue that is vague or seriously limited, and demonstrates weak critical thinking, providing inappropriate or insufficient examples, reasons, or other evidence to support its position is poorly organized and/or focused, or demonstrates serious problems with coherence or progression of ideas displays very little facility in the use of language, using very limited vocabulary or incorrect word choice demonstrates frequent problems in sentence structure contains errors in grammar, usage, and mechanics so serious that meaning is somewhat obscured.

**SCORE OF 1:** An essay in this category demonstrates very little or no mastery, and is severely flawed by ONE OR MORE of the following weaknesses: develops no viable point of view on the issue, or provides little or no evidence to support its position is disorganized or unfocused, resulting in a disjointed or incoherent essay displays fundamental errors in vocabulary demonstrates severe flaws in sentence structure contains pervasive errors in grammar, usage, or mechanics that persistently interfere with meaning.

Holistic score based on attached rubric (1–6): \_\_

## Appendix E. Sample Integrated Essays

*Global warming: Sample Essays scored 1 to 5*

### Sample Holistic Score 1

Climate Change is an important issue in the world we live in today. It changes some aspects of life drastically and is only a celerating with our constant mistreatment of our hone. We are responsible for the many cause of these changes due to.

The UN has concluded that since 1750 CO2 emissions have increase 31%, methane by 151%, and nitrogen oxide by 17%. I'm so so sorry, but i'm extremly tired. I can't even see straight, but I need this money. If the Artiiic disturbed we could experience a drop in temp by 3–5 degrees Celsius. We could possibly lose ice in greenland as well as the artic. I feel like i've been drugged. The ice melting ice could have conqueneces so large taht a passage opens uo from London to Tokyo. Several expersts hace predicted the impending natural anomlies that will occur.

Argumentation	Source Use	Language Sophistication	Organization	Holistic
1.5	1	2	1.5	1

This essay was scored as a 1 (very poor). The essay provides only one side of the argument (argumentation). The content and numbers provided in the essay indicate that the writer was referring to Source B (“Since pre-industrial times (around 1750) the concentration of carbon dioxide (CO2) has increased by around 31%, the concentration of methane (CH4) has increased by around 151% and the concentration of nitrogen oxide (N2O) has increased by around 17%.”). However, the information has not been transformed or connected to other ideas, nor does the writer make reference to the source (source use and inferencing). The essay goes off-topic into a stream-of-consciousness and then sharply returns back to the task. Even the sentences that are about the topic lack connectives (organization) and are difficult to understand due to multiple spelling and mechanical errors (language sophistication).

### Sample Holistic Score 2

So it seems there are some truths regarding global warming that aren't know to the common public. Global warming has it's ups, but doesn't in anyway over weighs its cons. Climate change on earth mainly involves the climate gas, CO2. And CO2 is a necessary gas in regards to its participation in the ecosystem such as photosynthesis and cellular respiration. However as it regards to the burning of fossil fuels, it results in “surplus carbon dioxide which remains in the atmosphere for a long time (Source B).” Leading to global warming and all its consequences.

The results of global warming could lead to the discovery of necessary resources such as petroleum, but the more catostrophic result of global warming is it's consequential increase in natural disasters such as hurricance's and floods. Humans play a crucial role in the formation of global warming. Due to their daily activities and the burning of fossil fuels.

Argumentation	Source Use	Language Sophistication	Organization	Holistic
2	2	2	2	2

This essay was scored as a 2 (poor). It shows a basic, but flawed command of argumentation, source use, language, and organization. This essay discussed two sides of the topic by stating the “ups” and “cons” of global warming, but does not address the prompt which asks about the extent to which global warming is a manmade phenomena until the penultimate sentence. The essay makes cites one source (Source B). The content suggests that the writer is also referring to Source C (e.g., hurricanes), but this is not made explicit by the writer. There is also no integration of the ideas from the two sources. Rather, the ideas from each source are “stacked”. The writer brings in some prior knowledge about CO2, but this information does not seem to be relevant to the issue at hand. The organization was somewhat logical with the help of rudimentary transition wordings (e.g., and, however). This essay shows fewer spelling and mechanical errors than the 1 essay, but still contains errors.

### Sample Holistic Score 3

In this day and age people all over the world are experiencing global warming, if they know it or not. The earth has changed dramatically overtime ever since the introduction to all sorts of CO2 producers, such as oil, gas, and methane and to further contribute to this phenomena known as global



warming, deforestation and/or desertification. In which each contributes to global warming as stated in source B and A. Frankly deforestation should be illegal and we should all help the earth and make peace but since these trees continue to be knocked down we face the effect of our actions and we continue to cut down trees. This causes negative consequences for the entire human race and the planet as a whole.

For example Mankind continues to have an effect of global warming due to various factors such as the pollution on the surface and in the water as well as burning of fossil fuels and other green house gasses. Green house gas is not natural to the earth as it is man made like fossil fuels and its causing the earth great harm. Harm caused is an example of desertification which makes land unusable and unable to yield any crops and can't be restored.

What is happening to the sun rays that is being reflected off of the earths surface and back to the atmosphere. What happens to the clouds and gases in the atmosphere after radiating the heat back to us and how does that effect us as humans. or even what would happen if the shield are the earth was broken beyond repair, then we would be burned to a crisp and human life would cease to ever exist along with all other forms of life.

Due to global warming rapidly changing the earth how would that change the heat shields being produced. is it a possibility that the green house gases are being compared to the earth's natural defense mechanism against the sun. These changes of global warming are being caused by us as people. with our greenhouse gases, methane and fossil fuels.

The carbon dioxide caused by human intervention is causing the significant increase in temperature. If this continues to go on and humans can't find an alternative to this problem then earth as we know it will cease to exist. Because if the temperature continues to rise then we will be unable to live on the planet.

Argumentation	Source Use	Language Sophistication	Organization	Holistic
3	2.5	3	3	3

This essay was scored as a 3 (fair). The essay cites two sources (A and B), but references them in passing without identifying what content comes from which source. In fact, the writer named several sources of greenhouse gases (e.g., oil, gas, and fossil fuels) from Source A and B and later paraphrased the mechanism behind greenhouse phenomenon from Source B (“What happens to the clouds and gases in the atmosphere after radiating the heat back to us and how does that effect us as humans. or even what would happen if the shield are the earth was broken beyond repair, ...”). This essay showed some critical thinking by discussing two sides of the topic with multiple relevant claims of global warming found in two sources. It also explicitly stated a position. The language of this essay suggested a developing logic with some degree of syntactic sophistication. It also included rudimentary transitions (e.g., for example, because) as well.

*Sample Holistic Score 4*

Climate change affects life on earth in numerous ways. Humans are responsible for many causes of climate change. As humans fail to realize the damage that is being done today, future generations won't live to see the green earth and blue waters. Are we meant to adapt, or are we going to be wiped out?

Climate change has always been around, but today, mankind has caused climate change to occur at a rapid and unnatural rate. According to Patrick Heskstad et al., carbon dioxide in the air continues to increase (Source A). Naturally, trees release carbon dioxide as they rot, and humans and animals release carbon dioxide as they breathe. However, humans have unnaturally increased the level of carbon dioxide in the atmosphere through the burning of fossil fuels, oil, gas, and coal (Source A and Source B). As a result of human actions, other climate gases have been released into the atmosphere, and the greenhouse effect increases, causing the temperatures to rise (Source A).

The effects of climate change on the rest of the earth is bizzare. Journalist Gustav Jensen explains how the oceans and weather changes. In their actions, mankind fails to acknowledge how the animals and plants are also affected. Homes and biomes are destroyed by deforestation and urbanization. Not only does it harm other organisms, but humans harm themselves in the long run. Jensen explains that a disruption in the Atlantic's circulation will cause a fall in temperature (Source C). Farming and forestry will be affected, while the need for heating increases (Source C). As the ice poles melt, ocean levels rise as the increase in evaporation occurs simultaneously (Source C). Developing countries, especially those in coastal communities, will be affected by intense hurricanes (Source C). More people will have to move inland.

Journalist John Hultgren tries to highlight the opprtunities that could be opened thorough climate change. He claims that riches could be accessed: oil, gas, gold, diamonds, copper, and zinc (Source D). But are they really riches if the rest of the world becomes lost in the process? He contradicts Jensen's statement concerning farming. Instead, he argues that the melting ice could allow for “better growing conditions and lower heating costs” (Source D). But what about cooling costs? Humans mst take responsibility now, or our descendants will suffer the consequences.

Argumentation	Source Use	Language Sophistication	Organization	Holistic
3.5	4	4	3	4

This essay was scored as a 4 (good). This essay demonstrated competent critical thinking by discussing two sides of the topic with evidence of global warming found in all four sources, which were integrated throughout. However, it implicitly stated their own position. The language of this essay exhibited a coherent logic, appropriate word choices and syntactic sophistication with a few instances of quality lapse.

*Sample Holistic Score 5*

Climate change is catastrophic for life on Earth and for the future of the generations ahead. And humans are primarily responsible for it. The growth of greenhouse gases in our atmosphere has led to rising sea levels, droughts, higher temperatures around the world, and more. People are being displaced in developing countries, hurricanes and floods are becoming more common place, and our glaciers are melting. This is a massive danger for our futures, and the damage will soon become irreversible.

The greenhouse effect is natural, and we are prone to panic when we hear of it, but it has actually been happening since the dawn of time. When climate gases are caught in the Earth and cannot radiate out of the atmosphere, we experience a warming experience throughout the planet (Source A).

This is not all bad, as if we did not have the greenhouse effect, we would be living at an average temperature of  $-18$  degrees Celsius instead of  $15$  degrees Celsius (Source A). However, research shows that humans are the reason why the greenhouse effect is getting worse and consistently having a worse impact for the Earth. Manmade carbon dioxide excretions are because of our use of coal, oil, and gas, as well as deforestation (Source B).

Our ecosystem is naturally regulating through natural selection, predator-prey relationships, viruses and growth, etc. And the balance that nature has is extremely fragile if tampered with. In fact some of the consequences of human fossil fuel use includes the melting of the Arctic as well as rising sea levels and the displacement of communities in developing countries (Source C).

Despite all of these consequences being something to fear and be worried about, this simply does not stop and will not stop big fossil fuel industries and capitalists. For them, this is another cash cow to build on. According to Source D, “The northerly regions that are becoming accessible also conceal enormous riches.” “Becoming accessible” is a very sugar-coated term for melting, dying, and causing disaster. For them, global warming will make industry better.

Climate change is caused by people, and people are feeling the effects of it. Our extensive use of fossil fuels since the industrial revolution has grown these indefeatable capitalist villains who now wish to profit off the detriment of others. Humans must save the Earth now before it will be too late to save it at all.

Argumentation	Source Use	Language Sophistication	Organization	Holistic
4	4	4	4	5

This essay was scored as a 5 because the participant displayed a reasonably consistent mastery of logic development, language, organization and source use. This essay was effective in developing an argument and thinking critically by coherently discussing two sides of the topic with multiple relevant claims of global warming found in all four sources. This essay demonstrated a good command of lexical and syntactic sophistication with a few lapses in quality. It included transition words (e.g., however, despite) to logically link sentences and paragraphs.

## References

- Alexander, P. A., Kulikowich, J. M., & Schulze, S. K. (1994). The influence of topic knowledge, domain knowledge, and interest on the comprehension of scientific exposition. *Learning and Individual Differences*, 6(4), 379–397. [https://doi.org/10.1016/1041-6080\(94\)90001-9](https://doi.org/10.1016/1041-6080(94)90001-9)
- Allen, L. K., Snow, E. L., Crossley, S. A., Jackson, G. T., & McNamara, D. S. (2014). Reading comprehension components and their relation to writing. *L'Année Psychologique*, 114(4), 663–691. <https://doi.org/10.4074/S0003503314004047>
- Allen, L. K., Snow, E. L., & McNamara, D. S. (2016). The narrative waltz: The role of flexibility on writing performance. *Journal of Educational Psychology*, 108, 911–924. <https://doi.org/10.1037/edu0000109>
- 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(5), 873–894. <https://doi.org/10.1007/s11251-013-9263-8>
- Barzilai, S., & Strømsø, H. I. (2018). Individual differences in multiple document comprehension. In J. L. G. Braasch, I. Bråten, & M. T. McCrudden (Eds.), *Handbook of multiple source use* (pp. 99–116). New York, NY: Routledge. <https://doi.org/10.4324/9781315627496-6>
- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2014). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1), 1–48. <https://doi.org/10.18637/jss.v067.i01>
- Braasch, J. L. G., & Bråten, I. (2017). The discrepancy-induced source comprehension (D-ISC) model: Basic assumptions and preliminary evidence. *Educational Psychologist*, 52(3), 167–181. <https://doi.org/10.1080/00461520.2017.1323219>
- Braasch, J. L., Bråten, I., & McCrudden, M. T. (Eds.). (2018). *Handbook of multiple source use*. New York, NY: Routledge. <https://doi.org/10.4324/9781315627496>
- Bråten, I., & Braasch, J. L. (2017). Key issues in research on students' critical reading and learning in the 21st century information society. In C. Ng, & B. Bartlett (Eds.), *Improving reading and reading engagement in the 21st century* (pp. 77–98). Singapore: Springer. [https://doi.org/10.1007/978-981-10-4331-4\\_4](https://doi.org/10.1007/978-981-10-4331-4_4)
- Bråten, I., Ferguson, L. E., Strømsø, H. I., & Anmarkrud, Ø. (2014). Students working with multiple conflicting documents on a scientific issue: Relations between epistemic cognition while reading and sourcing and argumentation in essays. *British Journal of Educational Psychology*, 84(1), 58–85. <https://doi.org/10.1111/bjep.12005>
- Bråten, I., Magliano, J. P., & Salmerón, L. (2020). Concurrent and task-specific self-reports. In D. L. Dinsmore, L. K. Fryer, & M. M. Parkinson (Eds.), *Handbook of strategies and strategic processing: Conceptualization, intervention, measurement, and analysis* (pp. 275–295). New York, NY: Routledge.
- Bråten, I., & Strømsø, H. I. (2010). When law students read multiple documents about global warming: Examining the role of topic-specific beliefs about the nature of knowledge and knowing. *Instructional Science*, 38(6), 635–657. <https://doi.org/10.1007/s11251-008-9091-4>
- Bråten, I., & Strømsø, H. I. (2011). Measuring strategic processing when students read multiple texts. *Metacognition and Learning*, 6(2), 111–130. <https://doi.org/10.1007/s11409-011-9075-7>
- Bråten, I., & Strømsø, H. I. (2020). On the roles of dispositions and beliefs in learning from multiple perspectives. In P. Van Meter, A. List, D. Lombardi, & P. Kendeou (Eds.), *Handbook of learning from multiple representations and perspectives* (pp. 141–163). New York, NY: Routledge.
- Britt, M. A., & Aglinskias, C. (2002). Improving students' ability to identify and use source information. *Cognition and Instruction*, 20(4), 485–522.
- 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, A. C. Graesser, & P. van den Broek (Eds.), *Narrative, comprehension, causality, and coherence: Essays in honor of Tom Trabasso* (pp. 209–233). Mahwah, NJ: Erlbaum.
- Britt, M. A., Rouet, J.-F., & Braasch, J. L. G. (2012). Documents as entities: Extending the situation model theory of comprehension. In M. A. Britt, S. R. Goldman, & J. F. Rouet (Eds.), *Reading: From words to multiple texts* (pp. 160–179). New York, NY: Routledge. <https://doi.org/10.4324/9780203131268>
- Britt, M. A., Rouet, J.-F., & Durik, A. M. (2017). *Literacy beyond text comprehension: A theory of purposeful reading*. New York, NY: Routledge. <https://doi.org/10.4324/9781315682860>
- Britt, M. A., Rouet, J.-F., & Durik, A. (2018). Representations and processes in multiple source use. In J. L. G. Braasch, I. Bråten, & M. T. McCrudden (Eds.), *Handbook of multiple source use* (pp. 17–33). New York, NY: Routledge. <https://doi.org/10.4324/9781315627496-2>
- van den Broek, P., Lorch, R. F., Linderholm, T., & Gustafson, M. (2001). The effects of readers' goals on inference generation and memory for texts. *Memory & Cognition*, 29(8), 1081–1087. <https://doi.org/10.3758/BF03206376>
- Chi, M. T., De Leeuw, N., Chiu, M. H., & LaVancher, C. (1994). Eliciting self-explanations improves understanding. *Cognitive Science*, 18(3), 439–477. [https://doi.org/10.1016/0364-0213\(94\)90016-7](https://doi.org/10.1016/0364-0213(94)90016-7)
- Coté, N., Goldman, S. R., & Saul, E. U. (1998). Students making sense of informational text: Relations between processing and representation. *Discourse Processes*, 25(1), 1–53. <https://doi.org/10.1080/01638539809545019>
- Cromley, J. G., & Azevedo, R. (2006). Self-report of reading comprehension strategies: What are we measuring? *Metacognition and Learning*, 1(3), 229–247. <https://doi.org/10.1007/s11409-006-9002-5>
- Cromley, J. G., & Azevedo, R. (2007). Testing and refining the direct and inferential mediation model of reading comprehension. *Journal of Educational Psychology*, 99(2), 311–325.
- Crossley, S. A., Roscoe, R., Graesser, A., & McNamara, D. S. (2011). Predicting human scores of essay quality using computational indices of linguistic and textual features. In G. Biswas, S. Bull, J. Kay, & A. Mitrovic (Eds.), *Proceedings of the 15th international conference on artificial intelligence in education* (pp. 438–440). Auckland, New Zealand: AIED. [https://doi.org/10.1007/978-3-642-21869-9\\_62](https://doi.org/10.1007/978-3-642-21869-9_62)
- Crossley, S. A., Weston, J. L., McLain Sullivan, S. T., & McNamara, D. S. (2011). The development of writing proficiency as a function of grade level: A linguistic analysis. *Written Communication*, 28(3), 282–311. <https://doi.org/10.1177/0741088311410188>
- 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.
- Dochy, F., Segers, M., & Buehl, M. M. (1999). The relation between assessment practices and outcomes of studies: The case of research on prior knowledge. *Review of Educational Research*, 69(2), 145–186.
- 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>
- Ferguson, L. E., Bråten, I., & Strømsø, H. I. (2012). Epistemic cognition when students read multiple documents containing conflicting scientific evidence: A think-aloud study. *Learning and Instruction*, 22(2), 103–120. <https://doi.org/10.1016/j.learninstruc.2011.08.002>

- Ferrari, M., Bouffard, T., & Rainville, L. (1998). What makes a good writer? Differences in good and poor writers' self-regulation of writing. *Instructional Science*, 26(6), 473–488. <https://doi.org/10.1023/A:1003202412203>
- Fitzgerald, J., & Shanahan, T. (2000). Reading and writing relations and their development. *Educational Psychologist*, 35(1), 39–50. [https://doi.org/10.1207/S15326985EP3501\\_5](https://doi.org/10.1207/S15326985EP3501_5)
- Flower, L., & Hayes, J. R. (1981). A cognitive process theory of writing. *College Composition & Communication*, 32(4), 365–387. <https://doi.org/10.2307/356600>
- Goldman, S. R. (2018). Discourse of learning and the learning of discourse. *Discourse Processes*, 55(5–6), 434–453. <https://doi.org/10.1080/0163853X.2018.1440866>
- Goldman, S. R., Britt, M. A., Brown, W., Cribb, G., George, M., Greenleaf, C., ... Project, R. E. A. D. I. (2016). Disciplinary literacies and learning to read for understanding: A conceptual framework for disciplinary literacy. *Educational Psychologist*, 51(2), 219–246. <https://doi.org/10.1080/00461520.2016.1168741>
- Goldman, S. R., & Scardamalia, M. (2013). Managing, understanding, applying, and creating knowledge in the information age: Next-generation challenges and opportunities. *Cognition and Instruction*, 31(2), 255–269. <https://doi.org/10.1080/10824669.2013.773217>
- Graham, S. (2018). A revised writer (s)-within-community model of writing. *Educational Psychologist*, 53(4), 258–279. <https://doi.org/10.1080/00461520.2018.1481406>
- Graham, S., & Hebert, M. (2011). Writing to read: A meta-analysis of the impact of writing and writing instruction on reading. *Harvard Educational Review*, 81(4), 710–744.
- Graham, S., Liu, X., Bartlett, B., Ng, C., Harris, K. R., Aitken, A., ... Talukdar, J. (2018). Reading for writing: A meta-analysis of the impact of reading interventions on writing. *Review of Educational Research*, 88(2), 243–284.
- Hammann, L. A., & Stevens, R. J. (2003). Instructional approaches to improving students' writing of compare-contrast essays: An experimental study. *Journal of Literacy Research*, 35(2), 731–756. [https://doi.org/10.1207/s15548430jlr3502\\_3](https://doi.org/10.1207/s15548430jlr3502_3)
- Haswell, R. H. (2000). Documenting improvement in college writing: A longitudinal approach. *Written Communication*, 17(3), 307–352. <https://doi.org/10.1177/0741088300017003001>
- Hayes, J. R. (1996). A new framework for understanding cognition and affect in writing. In C. M. Levy, & S. Randall (Eds.), *The science of writing: Theories, methods, individual differences, and applications* (pp. 1–27). Mahwah, NJ: Erlbaum.
- Hayes, J. R. (2012). Modeling and remodeling writing. *Written Communication*, 29(3), 369–388. <https://doi.org/10.1177/0741088312451260>
- Karimi, M. N. (2015). L2 multiple-document comprehension: Exploring the contributions of L1 reading ability and strategic processing. *System*, 52, 14–25. <https://doi.org/10.1016/j.system.2015.04.019>
- Kim, M., Tian, Y., & Crossley, S. A. (2021). Exploring the relationships among cognitive and linguistic resources, writing processes, and written products in second language writing. *Journal of Second Language Writing*, 53, Article 100824. <https://doi.org/10.1016/j.jslw.2021.100824>
- Kintsch, W. (1988). The role of knowledge in discourse comprehension: A construction-integration model. *Psychological Review*, 95(2), 163–182. <https://doi.org/10.1037/0033-295X.95.2.163>
- Kobrin, J. L., Deng, H., & Shaw, E. J. (2011). The association between SAT prompt characteristics, response features, and essay scores. *Assessing Writing*, 16(3), 154–169. <https://doi.org/10.1016/j.asw.2011.01.001>
- Kunda, Z. (1990). The case for motivated reasoning. *Psychological Bulletin*, 108(3), 480–498. <https://doi.org/10.1037/0033-2909.108.3.480>
- Kyle, K., & Crossley, S. (2016). The relationship between lexical sophistication and independent and source-based writing. *Journal of Second Language Writing*, 34, 12–24. <https://doi.org/10.1016/j.jslw.2016.10.003>
- Langer, J. A. (1984). Effects of topic knowledge on the quality and coherence of informational writing. In A. N. Applebee (Ed.), *Contexts for learning to write*. Norwood, NJ: Ablex.
- List, A., & Alexander, P. A. (2019). Toward an integrated framework of multiple text use. *Educational Psychologist*, 54(1), 20–39. <https://doi.org/10.1080/00461520.2018.1505514>
- List, A., Du, H., & Wang, Y. (2019). Understanding students' conceptions of task assignments. *Contemporary Educational Psychology*, 59, Article 101801. <https://doi.org/10.1016/j.cedpsych.2019.101801>
- MacGinitie, W. H. (1989). *Gates-MacGinitie reading tests*. Boston, MA: Houghton Mifflin.
- Magliano, J. P., McCrudden, M. T., Rouet, J.-F., & Sabatini, J. (2018). The modern reader: Should changes to how we read affect research and theory? In M. F. Schober, D. N. Rapp, & M. A. Britt (Eds.), *The Routledge handbook of discourse processes* (pp. 343–361). New York, NY: Routledge. <https://doi.org/10.4324/9781315687384-18>
- Maier, J., & Richter, T. (2016). Effects of text-belief consistency and reading task on the strategic validation of multiple texts. *European Journal of Psychology*, 31(4), 479–497. <https://doi.org/10.1007/s10212-015-0270-9>
- Martínez, I., Mateos, M., Martín, E., & Rijlaarsdam, G. (2015). Learning history by composing synthesis texts: Effects of an instructional programme on learning, reading and writing processes, and text quality. *Journal of Writing Research*, 7(2), 275–302. <https://doi.org/10.17239/jowr-2015.07.02.03>
- Mateos, M., Martín, E., Cuevas, I., Villalón, R., Martínez, I., & González-Lamas, J. (2018). Improving written argumentative synthesis by teaching the integration of conflicting information from multiple sources. *Cognition and Instruction*, 36(2), 119–138.
- Mateos, M., Solé, I., Martín, E., Cuevas, I., Miras, M., & Castells, N. (2014). In G. Rijlaarsdam, P. D. Klein, P. Boscolo, L. C. Kirkpatrick, & C. Gelati (Eds.), *Writing a synthesis from multiple sources as a learning activity* (pp. 169–190). Leiden: Brill. [https://doi.org/10.1163/9789004265011\\_009](https://doi.org/10.1163/9789004265011_009). Studies in writing, Vol. 28: Writing as a learning activity.
- McCarthy, K. S. (2020). Reading across the curriculum: Examining students' knowledge of disciplinary reading goals and strategies. In M. Gresalfi, & I. S. Horn (Eds.), 2. *The interdisciplinarity of the learning sciences, 14th international conference of the learning sciences (ICLS) 2020* (pp. 773–774). Nashville, Tennessee: International Society of the Learning Sciences. <https://doi.org/10.22318/icls2020.773>
- McCarthy, K. S., Kopp, K. J., Allen, L. K., & McNamara, D. S. (2018). Methods of studying text: Memory, comprehension, and learning. In H. Otani, & B. Schwartz (Eds.), *Research methods in human memory* (pp. 104–124). New York, NY: Routledge. <https://doi.org/10.4324/9780429439957-7>
- McCarthy, K. S., & McNamara, D. S. (2021). The multidimensional knowledge in text comprehension framework. *Educational Psychologist*, 56(3), 196–214. <https://doi.org/10.1080/00461520.2021.1872379>
- McCrudden, M. T., & Sparks, P. C. (2014). Exploring the effect of task instructions on topic beliefs and topic belief justifications: A mixed methods study. *Contemporary Educational Psychology*, 39(1), 1–11. <https://doi.org/10.1016/j.cedpsych.2013.10.001>
- McCutchen, D. (1986). Domain knowledge and linguistic knowledge in the development of writing ability. *Journal of Memory and Language*, 25(4), 431–444. [https://doi.org/10.1016/0749-596X\(86\)90036-7](https://doi.org/10.1016/0749-596X(86)90036-7)
- McCutchen, D. (2000). Knowledge, processing, and working memory: Implications for a theory of writing. *Educational Psychologist*, 35(1), 13–23.
- McCutchen, D., Covill, A., Hoyne, S. H., & Mildes, K. (1994). Individual differences in writing: Implications of translating fluency. *Journal of Educational Psychology*, 86(2), 256–266. <https://doi.org/10.1037/0022-0663.86.2.256>
- McNamara, D. S. (Ed.). (2007). *Reading comprehension strategies: Theories, interventions, and technologies*. Psychology Press.
- McNamara, D. S. (2011). Measuring deep, reflective comprehension and learning strategies: Challenges and successes. *Metacognition and Learning*, 3, 1–11.
- McNamara, D. S., & Allen, L. K. (2017). Toward an integrated perspective of writing as a discourse process. In M. Schober, A. Britt, & D. N. Rapp (Eds.), *Handbook of discourse processes* (2nd ed., pp. 362–389). New York, NY: Routledge. <https://doi.org/10.4324/9781315687384-19>
- McNamara, D. S., Crossley, S. A., & Roscoe, R. D. (2013). Natural language processing in an intelligent writing strategy tutoring system. *Behavior Research Methods*, 45, 499–515. <https://doi.org/10.3758/s13428-012-0258-1>
- McNamara, D. S., Levinstein, I. B., & Boonthum, C. (2004). iSTART: Interactive strategy training for active reading and thinking. *Behavior Research Methods, Instruments, & Computers*, 36(2), 222–233. <https://doi.org/10.3758/BF03195567>
- McNamara, D. S., Louwerse, M. M., McCarthy, P. M., & Graesser, A. C. (2010). Coh-Metrix: Capturing linguistic features of cohesion. *Discourse Processes*, 47, 292–330. <https://doi.org/10.1080/01638530902959943>
- McNamara, D. S., & Magliano, J. P. (2009). Towards a comprehensive model of comprehension. In B. Ross (Ed.), *The psychology of learning and motivation* (pp. 297–384). New York, NY: Elsevier. [https://doi.org/10.1016/S0079-7421\(09\)51009-2](https://doi.org/10.1016/S0079-7421(09)51009-2)
- McNamara, D. S., Roscoe, R., Allen, L., Balyan, R., & McCarthy, K. S. (2019). Literacy: From the perspective of text and discourse theory. *Journal of Language and Education*, 5(3), 56–69. <https://doi.org/10.17323/jle.2019.10196>
- Mokhtari, K., & Reichard, C. A. (2002). Assessing students' metacognitive awareness of reading strategies. *Journal of Educational Psychology*, 94(2), 249–259. <https://doi.org/10.1037/0022-0663.94.2.249>
- 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(2), 211–217.
- O'Reilly, T., & McNamara, D. S. (2007a). Reversing the reverse cohesion effect: Good texts can be better for strategic, high-knowledge readers. *Discourse Processes*, 43(2), 121–152. <https://doi.org/10.1080/01638530709336895>
- O'Reilly, T., & McNamara, D. S. (2007b). The impact of science knowledge, reading skill, and reading strategy knowledge on more traditional "high-stakes" measures of high school students' science achievement. *American Educational Research Journal*, 44(1), 161–196. <https://doi.org/10.3102/0002831206298171>
- van Ockenburg, L., van Weijen, D., & Rijlaarsdam, G. (2019). Learning to write synthesis texts: A review of intervention studies. *Journal of Writing Research*, 10(3), 401–428.
- OECD. (2019). *PISA 2018 assessment and analytical framework*. OECD Publishing. <https://doi.org/10.1787/b25efab8-en>
- Olinghouse, N. G., Graham, S., & Gillespie, A. (2015). The relationship of discourse and topic knowledge to fifth graders' writing performance. *Journal of Educational Psychology*, 107(2), 391–406. <https://doi.org/10.1037/a0037549>
- Ozuru, Y., Dempsey, K., & McNamara, D. S. (2009). Prior knowledge, reading skill, and text cohesion in the comprehension of science texts. *Learning and Instruction*, 19(3), 228–242. <https://doi.org/10.1016/j.learninstruc.2008.04.003>
- Perfetti, C. A., Rouet, J.-F., & Britt, M. A. (1999). Towards a theory of documents representation. In H. van Oostendorp, & S. R. Goldman (Eds.), *The construction of mental representations during reading* (pp. 99–122). Lawrence Erlbaum Associates, Inc.
- Primor, L., & Katzir, T. (2018). Measuring multiple text integration: A review. *Frontiers in Psychology*, 9, 2294. <https://doi.org/10.3389/fpsyg.2018.02294>
- Roscoe, R., Crossley, S. A., Snow, E. L., Varner, L. K., & McNamara, D. (2014). Writing quality, knowledge, and comprehension correlates of human and automated essay scoring. In 27th international Florida artificial intelligence research society conference, FLAIRS (pp. 393–398). The AAAI Press, 2014.
- Rouet, J.-F., Britt, M. A., & Durik, A. M. (2017). Resolv: Readers' representation of reading contexts and tasks. *Educational Psychologist*, 52(3), 200–215. <https://doi.org/10.1080/00461520.2017.1329015>
- Sabatini, J., O'Reilly, T., Weeks, J., & Wang, Z. (2019). Engineering a 21st Century reading comprehension assessment system utilizing scenario-based assessment techniques. *International Journal of Testing*. <https://doi.org/10.1080/15305058.2018.1551224>

- Saddler, B., & Graham, S. (2007). The relationship between writing knowledge and writing performance among more and less skilled writers. *Reading & Writing Quarterly*, 23(3), 231–247. <https://doi.org/10.1080/10573560701277575>
- Salmerón, L., Gil, L., Bråten, I., & Strømso, H. (2010). Comprehension effects of signaling relationships between documents in search engines. *Computers in Human Behavior*, 26(3), 419–426. <https://doi.org/10.1016/j.chb.2009.11.013>
- Sanchez, C. A., Wiley, J., & Goldman, S. R. (2006). Teaching students to evaluate source reliability during Internet research tasks. In S. A. Barab, K. E. Hay, & D. T. Hickey (Eds.), *Proceedings of the seventh international conference on the learning sciences* (pp. 662–666). Bloomington: International Society of the Learning Sciences.
- Scharrer, L., & Salmerón, L. (2016). Sourcing in the reading process: Introduction to the special issue. *Reading and Writing*, 29(8), 1539–1548. <https://doi.org/10.1007/s11145-016-9676-2>
- Spivey, N. N. (1997). The constructivist metaphor: Reading, writing and the making of meaning. *College Composition & Communication*, 48(4), 588–589.
- Spivey, N. N., & King, J. R. (1989). Readers as writers composing from sources. *Reading Research Quarterly*, 24(1), 7–26. <https://doi.org/10.1598/RRQ.24.1.1>
- Strømso, H. I., Bråten, I., & Britt, M. A. (2010). Reading multiple texts about climate change: The relationship between memory for sources and text comprehension. *Learning and Instruction*, 20(3), 192–204. <https://doi.org/10.1016/j.learninstruc.2009.02.001>
- Tierney, R. J., & Shanahan, T. (1991). Research on the reading–writing relationship: Interactions, transactions, and outcomes. In R. Barr, M. L. Kamil, P. B. Mosenthal, & P. D. Pearson (Eds.), *Vol. 2. Handbook of reading research* (pp. 246–280). Lawrence Erlbaum Associates, Inc.
- Van Meter, P., List, A., Lombardi, D., & Kendeou, P. (2020). *Handbook of learning from multiple representations and perspectives*. New York, NY: Routledge.
- Vandermeulen, N., van den Broek, B., Van Steendam, E., & Rijlaarsdam, G. (2020). In search of an effective source use pattern for writing argumentative and informative synthesis texts. *Reading and Writing*, 33(2), 239–266.
- Wang, Z., O'Reilly, T., Sabatini, J., McCarthy, K. S., & McNamara, D. S. (2021). A tale of two tests: The role of topic and general academic knowledge in traditional versus contemporary scenario-based reading. *Learning and Instruction*, 73, Article 101462. <https://doi.org/10.1016/j.learninstruc.2021.101462>
- Watanabe, M., McCarthy, K. S., & McNamara, D. S. (2019). Examining the effects of adaptive task selection on students' motivation in an intelligent tutoring system. In, *LAK'19. Proceedings of the 9th international conference on learning analytics and knowledge* (pp. 161–162).
- Wiley, J., Goldman, S. R., Graesser, A. C., Sanchez, C. A., Ash, I. K., & Hemmerich, J. A. (2009). Source evaluation, comprehension, and learning in Internet science inquiry tasks. *American Educational Research Journal*, 46(4), 1060–1106. <https://doi.org/10.3102/0002831209333183>
- Wiley, J., Griffin, T. D., Steffens, B., & Britt, M. A. (2020). Epistemic beliefs about the value of integrating information across multiple documents in history. *Learning and Instruction*, 65, Article 101266. <https://doi.org/10.1016/j.learninstruc.2019.101266>
- Wiley, J., Hastings, P., Blaum, D., Jaeger, A. J., Hughes, S., Wallace, P., ... Britt, M. A. (2017). Different approaches to assessing the quality of explanations following a multiple-document inquiry activity in science. *International Journal of Artificial Intelligence in Education*, 27(4), 758–790. <https://doi.org/10.1007/s40593-017-0138-z>
- Wiley, J., Jaeger, A. J., & Griffin, T. D. (2018). Effects of task instructions on comprehension from multiple sources in history and science. In J. L. G. Braasch, I. Bråten, & M. T. McCrudden (Eds.), *Handbook of multiple source use* (pp. 341–361). New York, NY: Routledge.
- Wiley, J., Steffens, B., Britt, M. A., & Griffin, T. D. (2014). Series Ed. In G. Rijlaarsdam, P. Klein, P. Boscolo, C. Gelati, & L. Kilpatrick (Eds.), *Writing to learn from multiple-source inquiry activities in history* (pp. 120–148). Leiden/Boston: Brill. [https://doi.org/10.1163/9789004265011\\_007](https://doi.org/10.1163/9789004265011_007) (Volume Eds.), *Studies in writing*, Vol. Vol. 28: Writing as a learning activity
- Wiley, J., & Voss, J. F. (1996). The effects of 'playing historian' on learning in history. *Applied Cognitive Psychology*, 10(7), 63–72. [https://doi.org/10.1002/\(SICI\)1099-0720\(199611\)10:7<63::AID-ACP438>3.0.CO;2-5](https://doi.org/10.1002/(SICI)1099-0720(199611)10:7<63::AID-ACP438>3.0.CO;2-5)
- 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.
- 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>
- Wissinger, D. R., De La Paz, S., & Jackson, C. (2021). The effects of historical reading and writing strategy instruction with fourth- through sixth-grade students. *Journal of Educational Psychology*, 113(1), 49–67. <https://doi.org/10.1037/edu0000463>
- 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](https://doi.org/10.1207/s1532690xci2304_2)