Optional ERIC Coversheet — Only for Use with U.S. Department of Education Grantee Submissions

This coversheet should be completed by grantees and added to the PDF of your submission if the information required in this form **is not included on the PDF to be submitted**.

INSTRUCTIONS
- Before beginning submission process, download this PDF coversheet if you will need to provide information not on the PDF.
- Fill in all fields—information in this form **must match** the information on the submitted PDF and add missing information.
- Attach completed coversheet to the PDF you will upload to ERIC [use Adobe Acrobat or other program to combine PDF files]—do not upload the coversheet as a separate document.
- Begin completing submission form at https://eric.ed.gov/submit/ and upload the full-text PDF with attached coversheet when indicated. Your full-text PDF will display in ERIC after the 12-month embargo period.

GRANTEE SUBMISSION REQUIRED FIELDS

Title of article, paper, or other content

All author name(s) and affiliations on PDF. If more than 6 names, ERIC will complete the list from the submitted PDF.

Last Name, First Name	Academic/Organizational Affiliation	ORCID ID

Publication/Completion Date—(if *In Press,* enter year accepted or completed)

Check type of content being submitted and complete one of the following in the box below:
- If article: Name of journal, volume, and issue number if available
- If paper: Name of conference, date of conference, and place of conference
- If book chapter: Title of book, page range, publisher name and location
- If book: Publisher name and location
- If dissertation: Name of institution, type of degree, and department granting degree

DOI or URL to published work (if available)

Acknowledgement of Funding— Grantees should check with their grant officer for the preferred wording to acknowledge funding. If the grant officer does not have a preference, grantees can use this suggested wording (adjust wording if multiple grants are to be acknowledged). Fill in Department of Education funding office, grant number, and name of grant recipient institution or organization.

"This work was supported by U.S. Department of Education **[Office name]** through **[Grant number]** to **Institution]** .The opinions expressed are those of the authors and do not represent views of the **[Office name]** or the U.S. Department of Education.

Integration in Multiple-Document Comprehension: A Natural Language Processing Approach

Allison N. Sonia, Joseph P. Magliano, Kathryn S. McCarthy, Sarah D. Creer, Danielle S. McNamara & Laura, K. Allen

Published online: 06 Jun 2022.

Submit your article to this journal ☞
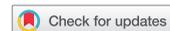
Article views: 348

View related articles ☞

View Crossmark data ☞

Integration in Multiple-Document Comprehension: A Natural Language Processing Approach

Allison N. Sonia[a], Joseph P. Magliano[b], Kathryn S. McCarthy[b], Sarah D. Creer[a], Danielle S. McNamara[c], and Laura, K. Allen[a]

[a]Department of Psychology, University of New Hampshire; [b]Department of Learning Sciences, Georgia State University; [c]Department of Psychology, Arizona State University

ABSTRACT

The constructed responses individuals generate while reading can provide insights into their coherence-building processes. The current study examined how the cohesion of constructed responses relates to performance on an integrated writing task. Participants ($N = 95$) completed a multiple document reading task wherein they were prompted to think aloud, self-explain, or evaluate the sources while reading and then write an integrated essay based on their reading. Natural Language Processing techniques were used to analyze the cohesion of the constructed responses at both within- and across-text levels. Both within- and across-text cohesion indices were positively related to essay quality; however, across-text cohesion indices exhibited stronger effects. Overall, this study provides evidence that the cohesion of constructed responses can serve as a proxy of the coherence of the mental representations that readers construct during multiple document processing.

Introduction

Comprehension is a complex process that relies on multiple subprocesses, from lower-level reading processes such as word recognition to strategies for *integrating* multiple levels of knowledge (e.g., vocabulary, domain) with text content (McNamara & Magliano, 2009). The majority of text comprehension models agree that the outcome of this process is a mental representation of the text and its meaning (e.g., Gernsbacher, 1990; Kintsch, 1998). This representation is said to be *coherent* to the extent that a reader establishes connections among the text information as well as to relevant prior knowledge (McNamara & Magliano, 2009). Thus, successful comprehension relies on creating, maintaining, and updating these connections while reading.

Importantly, much of the prior work in this area has focused on how readers process, comprehend, and remember *single* texts, despite the fact that readers often engage with multiple texts when they are attempting to learn new information (Magliano et al., 2018b; Snow, 2002). The ability to comprehend information from *multiple* texts has become a critical skill in the increasingly digital era (Goldman et al., 2012; Braasch et al., 2018; List & Alexander, 2019; Magliano et al., 2018b; Rouet, 2006). Readers today have relatively quick and easy access to information from a variety of sources. This may present comprehension difficulties, as readers are challenged to integrate across sources, which can vary in reliability or may include contradictory information (Britt & Rouet, 2012; Britt et al., 2013; Goldman, 2004; Rouet & Britt, 2011).

CONTACT Allison N. Sonia ✉ allison.sonia@unh.edu 🖷 Psychology Department, University of New Hampshire, 468 McConnell Hall, 15 Academic Way, Durham, NH 03824

This article has been corrected with minor changes. These changes do not impact the academic content of the article.

In response to these new challenges, there has been a growing interest in examining how readers comprehend multiple documents (MDs), particularly when they include a mixture of reliable and unreliable sources or contain conflicting or inaccurate information (e.g., Barzilai & Strømsø, 2018; G. Braasch et al., 2012; Stadtler & Bromme, 2014). In this MD context, readers must not only have sufficient knowledge and strategies to comprehend each document individually, but also to *evaluate* the quality of the sources, *integrate* the information presented across the documents, and *use* this information to ideally provide a well-reasoned argument (Braasch et al., 2018). Despite the evident differences in MD comprehension, it has been commonly assumed that the basic processes supporting single text comprehension go relatively unchanged in MD contexts; however, these assumptions have not been tested directly (Britt et al., 2013; Goldman et al., 2013; List & Alexander, 2019; Saux et al., 2021). Research has therefore predominantly focused on examining the processes involved in evaluating sources rather than examining how basic comprehension processes unfold in MD reading contexts. It is likely the case that many of the basic processes involved in single text comprehension are also relevant to MD comprehension; however, these assumptions need to be empirically tested.

The purpose of the current study was to examine the integration processes underlying MD comprehension by examining participants' constructed responses to texts while reading. Participants were instructed to provide one of three types of constructed responses while reading multiple texts: think-alouds, self-explanations, or source evaluations; they then wrote an integrated essay based on these texts. The cohesion (i.e., the overlap between response segments in terms of words and concepts) of these three types of responses was analyzed using Natural Language Processing (NLP) techniques to provide a proxy for whether participants integrated information across these documents while reading. We aimed to examine how participants' online reading processes (as assessed via the cohesion of these constructed responses) were related to the quality of the essay they produced after reading. We first provide an overview of how constructed responses generated during reading can be used to measure connections made within and across documents followed by a brief discussion of the previous work investigating the outcomes of comprehension using integrated essays. The focus of this article is extending these methods of investigation to a MD reading context.

Constructed responses during reading

Prompting readers to generate *constructed responses* as they read is a common way of studying the cognitive processes that occur during comprehension (McCarthy et al., 2018; McNamara et al., 2006). Think-alouds are perhaps the most well-established procedure for collecting readers' thoughts during reading (Pressley & Afflerbach, 1995). When thinking aloud, readers are intermittently interrupted and asked to report their thoughts as they come to mind; as such, these instructions are relatively neutral in that they do not bias readers to adopt particular strategies (Ericsson & Simon, 1993). Readers' think-alouds are sensitive to changes in text structure as well as differences in their prior knowledge and metacognitive states while reading and therefore provide insights into coherence-building processes (Magliano & Millis, 2003; McNamara & Magliano, 2009). Think-aloud studies have aided in the identification of processes, such as inference generation, involved in comprehension across a variety of contexts (Magliano et al., 2020). For example, when compared to their less-skilled peers, skilled readers tend to not only generate more inferences but also tend to make more global connections (e.g., at the paragraph level) rather than only local connections (e.g., at the level of adjacent sentences; Karlsson et al., 2018; Millis et al., 2006; Wolfe & Goldman, 2005).

Beyond simply thinking aloud, the instructions that readers receive for generating constructed responses can be used to provoke specific strategic processes while reading (e.g., Allen, McNamara, & McCrudden, 2015; Chi et al., 1994). These strategy instructions can encourage students to attend to specific aspects of a text and potentially increase their performance on a subsequent task. The current study used two different strategy instructions to investigate their impact on MD comprehension. The first strategy we investigated was source evaluation. Several theoretical frameworks of MD comprehension have highlighted the importance of *sourcing* as a critical process for comprehending MDs

(e.g., the Documents Model Framework; Britt et al., 1999, 2013; Britt & Rouet, 2012; Perfetti et al., 1999; Rouet, 2006; MD-TRACE; Rouet & Britt, 2011; D-ISK; Braasch & Bråten, 2017). *Sourcing* refers to the readers' ability to attend to and evaluate source features such as the author(s), their credentials, and the publisher (Bråten, Strømsø, & Anne, 2009; Britt & Aglinskas, 2002; Rouet & Britt, 2011). Source features are seen as important for MD processing because there can be dramatic differences in the reliability of the sources that might be encountered in an MD task. Several theories of MD comprehension argue that source features are included in the mental representation of the texts and that connections to source information allow readers to better integrate textual information (Anmarkrud et al., 2014; Bråten & Strømsø, 2012; Britt & Aglinskas, 2002; Britt et al., 1999; Goldman, 2004; Rouet et al., 2009).

Integral to the processes of sourcing in MD comprehension is a reader's ability to strategically evaluate the credibility of a given source. Credibility is often characterized as a combination of two primary components: expertise (i.e., the qualifications of the source/author) and trustworthiness (i.e., the ability of the source to provide true and unbiased information; Hovland et al., 1953). A reader's ability to judge source credibility is often linked to their comprehension ability (Anmarkrud et al., 2014; Barzilai et al., 2015; Bråten et al., 2009; Goldman et al., 2012). However, prior research indicates that even skilled readers often fail to critically evaluate document credibility while conducting MD tasks (Britt & Aglinskas, 2002; Claassen, 2012; Wineburg, 1991). Source evaluation has been considered critically important in the modern internet era, wherein information can be shared and circulated regardless of its accuracy (Magliano et al., 2018b).

Accordingly, interventions in MD contexts have commonly focused on increasing readers' attention to source information through either *instruction to engage in sourcing* or through more extensive *source evaluation training* (Braasch, Bråten, Strømsø, Anmarkrud, & Ferguson, 2013; Britt & Aglinskas, 2002; Wiley et al., 2009). For example, Britt and Aglinskas (2002) provided secondary students with computer-based tutoring on strategies for analyzing source features of texts. The tutoring program trained readers to critically evaluate source information by asking them to provide written reflections on the features of the sources provided by the researchers. When compared to a control group that wrote reflections without guidance on source evaluation, students who received the training cited more sources in their responses and answered more questions about source information correctly in a posttest. Other interventions for MD tasks simply prompt readers to attend to and consider source information (Braasch, Bråten, Strømsø, Anmarkrud, & Ferguson, 2013; Stadtler et al., 2015; Wiley et al., 2009). However, while the outcomes of instructing readers to attend to source information are promising, the mechanisms behind *how* they support learning outcomes requires additional exploration. Investigating how instructional prompts lead to differences in comprehension during reading is critical to gaining further insight into the strategies that can most effectively support MD comprehension.

The second strategy that was investigated in the current study is self-explanation. Self-explanation involves monitoring your understanding and explaining the text to yourself as you read. Prompting students to self-explain increases inference generation and connections to prior knowledge, thus supporting deeper comprehension (McNamara, 2004). To date, self-explanation has predominantly been applied in single-document comprehension but may also provide a means for supporting MD comprehension. MD tasks require readers to represent important relationships conveyed both within individual texts and across multiple texts in a set (Rouet & Britt, 2011). Establishing connections and integrating information across multiple texts is a particularly difficult task as there are more distal connections between ideas and fewer discourse markers to support inference generation than in single document scenarios (Britt & Rouet, 2012; Goldman & Rakestraw, 2000; Rouet, 2006; Saux et al., 2021). Therefore, readers must develop goals and strategies that promote integration across texts, particularly when the task, such as writing an essay, requires integration across multiple texts (Rouet & Britt, 2011).

Given the success of self-explanation in promoting the integration of information within single texts, it follows that this strategy has strong potential to support integration in an MD context. Preliminary evidence supports this extension: think-aloud protocols collected during MD tasks have

demonstrated that skilled readers are more likely to spontaneously engage in self-explanation strategies than less-skilled readers (Anmarkrud et al., 2014; Goldman et al., 2012; Wolfe & Goldman, 2005). Despite this evidence, there have been few investigations into the extent to which self-explanation promotes successful comprehension of MDs via an increase in connections generated *within* and *across text*. In the context of MD comprehension, collecting constructed responses based on think-aloud, self-explanation, and source evaluation instructions has strong potential to provide valuable insights into how readers process multiple texts and the strategies that best support successful comprehension.

Measuring cohesion of constructed responses

The *coherence* of a reader's mental representation is important for their deep comprehension of texts (see McNamara & Magliano, 2009). Coherence is established through the activation and integration of prior knowledge and text content, which helps to create meaningful connections across concepts and ultimately construct meaning. However, *coherence* itself cannot be directly measured but instead must be inferred from comprehension assessments or computational models that simulate comprehension processes (e.g., Kintsch, 1998). Notably, most coherence-building processes occur *in the moment* as the reader progresses through the text (Kintsch, 1988). However, most assessments of text comprehension (as with integrated essay writing) occur after the reading task has finished, rendering it difficult to measure coherence-building processes (Magliano et al., 2007).

Examining constructed responses that readers generate during reading can provide an avenue for examining the online processes of coherence-building that potentially support comprehension. Research has turned to NLP methodologies to examine the properties of these constructed responses in a variety of discourse comprehension and production contexts, such as essays, social media posts, and narratives (Graesser & McNamara, 2012; Landauer et al., 2007; Shermis et al., 2010; for a review, see Allen et al., 2021). These NLP techniques allow computers to analyze language across multiple dimensions, which can provide researchers with more nuanced and fine-grained assessments of learning processes. For example, NLP can be used to calculate basic features of the text, such as its length; more complex information, such as the structure of the sentences; and provide information about its emotional tone and semantic flow.

One useful NLP metric for examining integration processes during reading is *cohesion*. Cohesion refers to the explicit cues in text that establish connections among text content (Gernsbacher, 1990; McNamara et al., 2014). For example, the repetition of words in a text, overlapping ideas within a text, and the use of connectives (e.g., "and," "because") are all markers of cohesion and indicate the presence of interconnected ideas. It has been established that the within-text cohesion (i.e., connections made within a single text) of readers' constructed responses during reading is indicative of the coherence of their mental representations (Allen et al., 2015). For example, Allen et al. (2015) used NLP to analyze the cohesion within students' constructed responses (i.e., the lexical and semantic overlap from one constructed response to the next) generated while reading single texts. They found that the cohesion of the constructed responses was higher when readers were prompted to self-explain compared to when readers were asked to paraphrase. Additionally, the cohesion of readers' constructed responses increased over the course of the self-explanation training, indicating that the readers were making more connections between ideas within the texts as they read. Thus, cohesion can provide a measure of coherence-building processes during comprehension; however, this hypothesis has yet to be tested in an MD comprehension context.

In MD contexts, readers need to not only maintain coherence for information within the individual texts but also across the different texts. Such coherence-building processes may be more challenging because the connections across texts are likely to be less explicit (Goldman & Rakestraw, 2000; Magliano et al., 2018b). Assessing across-text cohesion in readers' constructed responses can reveal

insights into MD-specific processes. Thus, measuring cohesion in an MD context can be used to model both connections made while thinking aloud for an individual text (i.e., within-text cohesion) and connections made across multiple texts (i.e., across-text cohesion).

Both of these cohesion measures have the potential to provide information about how readers build coherence while reading from multiple texts. Much like previous research in single text comprehension, cohesion for individual texts might examine the number of overlapping words and concepts between constructed responses written while reading a single text to provide insights into how the reader establishes an understanding of a single text in the text set. Alternatively, across-text cohesion can provide information about how readers are integrating information across all of the texts in the set. Thus, both measures are critical to understanding the coherence-building processes that support MD comprehension.

Integrated essays

In addition to looking at readers' constructed responses during reading, other methods of measuring MD comprehension can be used to investigate the outcome of the comprehension process. In MD studies, comprehension is commonly assessed by having readers produce integrated essays to demonstrate their understanding of the concepts in the texts (e.g., Anmarkrud et al., 2014; Britt & Aglinskas 2002; McCarthy et al., 2018; Rouet et al., 1996; Weston-Sementelli et al., 2016; Wiley et al., 2009). Integrated essay tasks often prompt participants to read a set of texts and then compose an argumentative essay integrating information from the sources that they have just read (Barzilai et al., 2015, 2015; List et al., 2019; McCarthy et al., 2018). These prompts are typically designed to reflect comprehension tasks that are common in both educational settings and the workplace (e.g., see Appendices A and B for the essay prompts and texts used in this study).

Traditionally, research using integrated essays as an outcome measure has focused on how readers select, source, and integrate information from the texts that they read (Braasch et al., 2018; Braasch et al., 2013, 2018; Wiley et al., 2020). For example, less-skilled readers tend to construct integrated essays that are either over-compartmentalized (e.g., abstract stacking) or too generalized (i.e., a "mush model"; Britt et al., 1999). By contrast, higher-quality integrated essays tend to be well organized in ways that reflect the integration of ideas spread across the document set (Britt & Rouet, 2012). In these studies, essays are typically evaluated based on correctness, number of source mentions, or quality of source integration. More recent work has argued a need to consider the contribution of writing skills to the quality of integrated essay writing (McNamara & Allen, 2017; McNamara et al., 2019). The generation of a high-quality integrated essay requires the development of both reading comprehension skills as well as writing skills related to the composition and organization of ideas (Weston-Sementelli et al., 2016). In the current study, integrated essays were used as a post-reading assessment of readers' comprehension of information from the multiple sources provided. We looked at multiple levels of writing quality of the essays; holistic (i.e., overall) quality, language sophistication, organization, argumentation, and source use.

Current study

The purpose of the current study was to examine the extent to which the cohesion detected in readers' constructed responses was predictive of argumentative, integrated essay quality. Here, we adopt a methodological approach that relies on theoretically motivated NLP techniques wherein we use the language that participants produce as they read (i.e., in their constructed responses) to understand how they are establishing connections within and across the texts in an MD reading and writing task. In particular, we manipulated the instructions that participants were given before reading a set of texts. Participants were instructed to *think aloud, self-explain*, or *evaluate the source material* as they read. We then calculated linguistic indices that we hypothesized would be indicative of coherence-building both within- and across-texts.

Natural language exhibits cohesion in a variety of different ways. Here, we focus on one form of cohesion: lexical cohesion (Crossley et al., 2019). *Lexical cohesion* refers to cohesion that manifests in the form of overlapping words across a text. That is, if a participant's constructed responses often repeat the same (or semantically similar) words, then the participant is demonstrating high lexical cohesion across their responses.

Importantly, prior studies have shown that within-text cohesion of constructed responses is predictive of comprehension and reading skill while reading single texts (Allen et al., 2015, p. 2016); however, this finding has yet to be generalized to an MD context. In the current study, we calculate lexical cohesion not only at the individual text level (*within-text* cohesion) but also across an entire text set (*across-text* cohesion). Thus, we examine the extent to which both within- and across-text lexical cohesion manifests in readers' constructed responses and the extent to which cohesion relates to the quality of essays composed after the reading task. We expect that in the context of MD comprehension, the connections captured by across-text cohesion will contribute more to essay quality than within-text cohesion. Finally, we examine whether the relations between cohesion and essay quality are moderated by instructional condition (i.e., think-aloud, self-explanation, and source evaluation).

Thus, we aimed to answer two primary research questions. The more focal question was *How do the within- and across-text cohesion indices in participants' constructed responses relate to the overall quality of participants' post-reading integrated essays?* We also manipulated reading instructions in order to generate a variety of types of processes in which readers engage during both single- and multiple-text comprehension tasks. These different strategies have been shown to differentially impact comprehension (e.g., Allen et al., 2015; Britt & Aglinskas, 2002). Thus, a second research question was *To what extent do different comprehension strategies (i.e., constructed response instructions) influence the relations between the cohesion of constructed responses and integrated essay writing quality?*

Methods

Participants

Ninety-five participants included 51 high school students (M_{age} = 16.22, SD_{age} = .96) who participated in the study in the fall of 2019 and 45 college freshmen students who had graduated high school the previous spring (M_{age} = 18.24, SD_{age} = .53) and participated in the fall of 2020. Data for two high school participants were removed from the analyses: one for missing days of data collection and one for lost data due to a technical error. One college participant's data was removed due to an incomplete demographic questionnaire, leaving 93 participants in the final analysis.

The sample was predominantly female, with 76 participants identifying as female and 17 participants identifying as male. The participant sample also predominantly identified as 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 participants (n = 82) were native English speakers. There were no significant differences in any of the dependent variables across gender or English as a native language status.

Materials and measures

Constructed response instructions

Participants were prompted to generate constructed responses periodically while reading. These prompts appeared six to nine times per document at preselected sentences, such that all participants responded at the same sentences. Participants were assigned to one of three constructed response instruction conditions: think-aloud, self-explanation, or source evaluation. The prompts provided to participants in each condition are presented in Appendix A. Participants in the think-aloud

condition were asked to report the thoughts that immediately came to mind regarding the text as they were reading. Participants in the self-explain condition were asked to provide their self-explanations of the text while they read. They were further instructed to explain the meaning of the text and to elaborate beyond their initial understanding. Participants in the source evaluation condition were asked to reflect on information about the source (i.e., author, publication, date/locations, audience) of the text while they read and to report their thoughts on how the source impacts the meaning of the text. In each condition, the instructions provided examples of each type of constructed response to a short text segment to give the participant a better idea of how their responses should be structured.

Document sets

The document sets read by the students were adapted from previous studies exploring MD processing (Anmarkrud et al., 2013; Ferguson et al., 2012; Strømsø et al., 2010). Each document set contained four texts; one set was focused on global warming and the other was focused on cell phone use. The presentation of the texts was counterbalanced and randomly assigned. In the context of the larger study (see McCarthy et al., 2022), our intent was to include a text set that contained few connections and one that contained an abundance of connections. However, in the current study, wherein we are attempting to examine the connections between constructed responses, the overlapping ideas in the cell phone use document set overwhelmed the ability to observe variation in readers' self-generated connections between ideas. In essence, the high cohesion between the documents comprising the cell phone text set results in a situation described by Magliano et al. (2018a), in which both human and computational scoring of essays and constructed responses is compromised because it is difficult, if not impossible, to determine where thoughts are coming from with respect to the text sets. As such, we restricted the analyses in this study to the global warming text set. For the global warming set, two texts discussed whether the causes of global warming were natural or manmade and two discussed the negative and positive consequences of global warming (Appendix B).

Integrated essay

After reading the document set, readers were asked to write an integrated essay that explained the effects of climate change for life on earth and the extent to which humans are responsible. Participants were encouraged to elaborate on the information in the text rather than summarizing it. They were also asked to use information from the texts to support their ideas, but to put ideas in their own words.

Procedure

This study took place over a series of experimental sessions which were completed in 1 week. Participants completed this study as part of a larger investigation of MD reading and writing using the MD module of Interactive Strategy Training for Active Reading and Thinking (iSTART; McNamara et al., 2004). During the session relevant to the current analyses, participants completed an MD task in iSTART that involved reading four texts on the topic of global warming. Participants were asked to skim the texts before engaging in the deeper reading process. During the reading portion of the experiment, participants were prompted to engage in one of three types of constructed response protocols: *think-aloud* ($n = 30$), *self-explain* ($n = 32$), or *evaluate the sources* ($n = 31$). They were then given 25 minutes to write an integrated essay.

Analyses

Essay scores

The essays were evaluated using a scoring rubric that included one holistic essay quality score and four subscale scores (i.e., argumentation, source use, language sophistication, and organization). This rubric was developed to tap into the quality of the writing at multiple dimensions (e.g., Crossley et al., 2021). The holistic essay scores ranged from 1 to 6 based on the following criteria: 1 = very poor, 2 = poor, 3 = fair, 4 = good, 5 = very good, 6 = excellent. The four subscale scores ranged from 1 to 4 (see the subscale rubric in Appendix C). *Argumentation* scores were based on how well the participants discussed the sides of the arguments presented in the texts and stated their position. *Source use* scores were based on the participants' reference of source information and interpretation of source material beyond simple paraphrasing. Scores of *language sophistication* were graded based on the level of sophistication and variety of participants' word choices as well as their ability to write without many grammar and spelling errors. Finally, *organization* scores were awarded based on the degree to which the essays were well structured.

Human ratings for the scores were provided by two teams of two expert raters each. The raters were PhD students in an English composition program, who had over 3 years of experience teaching writing at the university level and rating experience with standardized rubrics. The raters were first trained on the rubric using essays that were not part of the corpus used in this study. When raters reached an acceptable level of reliability ($\kappa > 0.70$), they scored the integrated essays in the current study such that two raters scored each essay. After initial scoring, kappa scores ranged from .48 to .65. Because this scoring was done at the essay level rather than the idea-unit level, there is a higher level of variability. Due to this variability in scores, raters adjudicated their scores for all essays together. If any differences in scores were greater than 1, raters discussed the scores and made adjustments as needed. After adjudication scores from the two raters were averaged, all kappa scores were greater than .6.

Automated cohesion analyses of the constructed responses

Both within- and across-text cohesion of the constructed responses was calculated using the Tool for the Automated Analysis of Cohesion (Crossley et al., 2019). For the purposes of the current study, we selected four indices and then calculated them both at within- and across-text levels (Figure 1). All cohesion indices were calculated at the participant level such that they reflected the average cohesion score for each participant. The four indices that we selected were intended to tap into coherence-building processes at multiple levels of analysis. First, we selected two lexical overlap indices that measured simply which words overlapped across participants' think-alouds. These indices were intended to establish when the participants were making connections among explicitly similar text content rather than simply talking about related ideas. These lexical overlap indices also were measured at two grain sizes, such that we could examine whether there were differences in integration
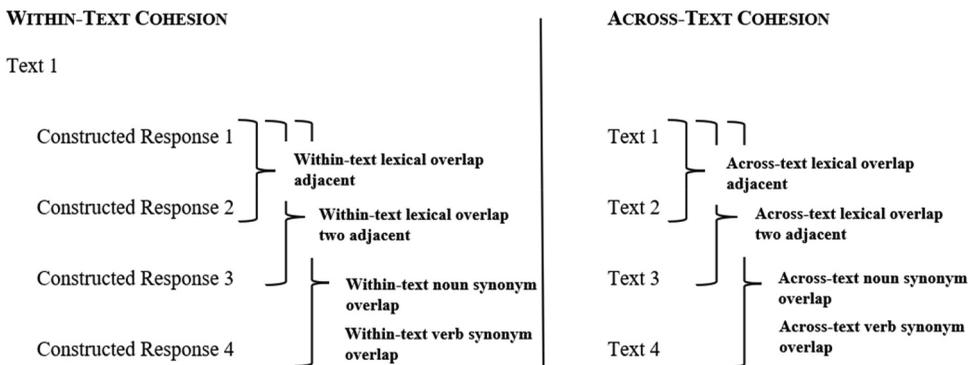

Figure 1. Measures of cohesion calculated within- and across-text.

at more local and global levels. Beyond these lexical overlap indices, we also measured semantic cohesion by examining overlap of noun and verb synonyms. These indices were intended to tap into more conceptual integration of concepts (nouns) and relations (verbs). There are a number of other ways in which cohesion can be calculated; however, due to our sample size, we aimed to keep the number of analyses small to avoid the risk of Type II errors. Therefore, we selected indices that would capture the primary dimensions at which we hypothesized cohesion would most strongly vary (i.e., explicit vs semantic and nouns vs verbs), which we anticipated would provide a strong foundation for future analyses.

Lexical overlap (adjacent). Within-text lexical overlap (adjacent) was measured as the average amount of explicit overlap in participants' words across adjacent constructed responses within each text. To calculate across-text indices the constructed responses for each text were treated as one singular response. Across-text lexical overlap (adjacent) was measured as the average amount of explicit overlap in participants' words across the constructed responses generated for each adjacent text. For each level, the indices were normed by the total number of responses (see Crossley et al., 2019).

Lexical overlap (two adjacent). Within- and across-text lexical overlap (two adjacent) was calculated in the same way as lexical overlap (adjacent) except that it included the amount of word overlap across two adjacent responses (within-text level) or across two adjacent texts (across-text level).

Noun synonym overlap. Within-text noun synonym overlap was calculated as the extent to which participants generated connections using semantically similar nouns within the responses generated for each text. Across-text noun synonym overlap was calculated as the extent to which participants generated connections using semantically similar nouns across the texts (when the responses for each text were treated as one response).

Verb synonym overlap. Within-text noun synonym overlap was calculated as the extent to which participants generated connections using semantically similar verbs within the responses generated for each text. Across-text noun synonym overlap was calculated as the extent to which participants generated connections using semantically similar verbs across the texts (when the responses for each text were treated as one response).

Results

Table 1 provides the descriptive statistics and correlations between the number of words in the essays and the five essay quality scores (i.e., one holistic essay quality score and four subscale scores). Participants' constructed responses contained a mean of 331.46 words (SD = 114.41, range = 148–670). The holistic essay score was correlated with all subscale scores, indicating that the four dimensions of essay quality were all related to the overall quality judgments of the raters. Importantly, there

Table 1. Descriptive statistics and correlations for essay scores (N = 93).

Variable	Mean	SD	Range	1	2	3	4	5
1. Number of words	331.46	114.41	148–670	–				
2. Argumentation	2.70	0.71	1–4	.53*	–			
3. Source use	2.60	0.78	1–4	.44*	.62*	–		
4. Language sophistication	2.83	0.51	2–4	.40*	.54*	.45*	–	
5. Organization	2.64	0.69	1–4	.59*	.69*	.59*	.51*	–
6. Holistic score	2.82	0.95	1–5	.63*	.80*	.78*	.62*	.81*

*p < .001

was no effect of constructed response instruction manipulation on overall essay score, $F(2,90) = .01$, $p = .99$ (think-aloud: $M = 2.82$, $SD = 0.92$; self-explanation: $M = 2.84$, $SD = 0.87$; source evaluation: $M = 2.81$, $SD = 1.08$).

Holistic essay quality and cohesion

Our first research question focused on the extent to which within- and across-text cohesion of participants' constructed responses was related to the overall quality of their essays. See Table 2 for the descriptive statistics for all cohesion indices. At the within-document level, holistic essay scores were significantly correlated with the lexical overlap indices but not the synonym indices (Table 3). However, all four of the across-text indices were related to holistic essay scores. These results suggest that connections made within and across documents in the document set were related to higher-quality integrated essays. However, the absence of a correlation between the within-document synonym indices and essay quality suggests that the coherence-building processes engaged while reading single texts may not translate as robustly to performance on integrated essays, particularly if those connections are only semantically related, rather than explicit.

We conducted a linear model to predict essay scores from the within- and across-text cohesion indices. For this model, we selected only the within- and across-text cohesion indices that were correlated with holistic essay scores. Two indices of within-text cohesion and four indices of across-text cohesion were correlated with holistic essay score and used as independent variables in the model. Specifically, the two indices of within-document cohesion were lexical overlap (adjacent) and lexical overlap (two-adjacent). The four indices of across-text cohesion were lexical overlap (adjacent), lexical overlap (two adjacent), noun synonym overlap, and verb synonym overlap. The dependent variable was holistic essay score. This analysis yielded a significant model that accounted for approximately 26% of the variance in essay score, $F(6, 86) = 4.879$, $p < .001$, $R^2 = .25$. Only one of the variables was a significant predictor in the model: across-text lexical paragraph overlap (two adjacent; $B = 0.06$, $t(6, 86) = 2.29$, $p < .05$).

Table 2. Descriptive statistics for cohesion indices (overall and by instructional condition).

Index	Within Mean (SD)	Across Mean (SD)
Overall (N = 93)		
Lexical overlap (adjacent)	5.53 (3.91)	23.83 (10.32)
Lexical overlap (two adjacent)	7.65 (4.42)	32.97 (14.09)
Noun synonym overlap	1.82 (2.57)	13.23 (10.76)
Verb synonym overlap	1.05 (1.12)	17.51 (14.64)
Think-aloud (n = 30)		
Lexical overlap (adjacent)	3.43 (1.72)	20.54 (9.22)
Lexical overlap (two adjacent)	5.10 (2.33)	28.70 (12.65)
Noun synonym overlap	0.64 (0.64)	6.06 (4.80)
Verb synonym overlap	0.60 (0.46)	12.75 (10.37)
Self-explain (n = 32)		
Lexical overlap (adjacent)	7.41 (5.23)	27.30 (10.59)
Lexical overlap (two adjacent)	9.83 (5.35)	37.62 (14.44)
Noun synonym overlap	2.94 (3.91)	18.31 (12.42)
Verb synonym overlap	1.51 (1.52)	22.96 (16.11)
Source evaluation (n = 31)		
Lexical overlap (adjacent)	5.62 (2.76)	23.42 (10.24)
Lexical overlap (two adjacent)	7.86 (3.68)	32.30 (14.02)
Noun synonym overlap	1.80 (1.16)	14.93 (9.55)
Verb synonym overlap	1.02 (0.91)	16.48 (15.14)

Table 3. Correlations between lexical cohesion and holistic essay quality (overall and by instructional condition.

Index	Within	Across
Overall (N = 93)		
Lexical paragraph overlap (adjacent)	0.21*	0.42**
Lexical paragraph overlap (two adjacent)	0.25*	0.46**
Noun synonym overlap	0.15	0.28**
Verb synonym overlap	0.17	0.31**
Think-aloud (n = 30)		
Lexical paragraph overlap (adjacent)	0.07	0.26*
Lexical paragraph overlap (two adjacent)	0.11	0.26*
Noun synonym overlap	0.18	0.16
Verb synonym overlap	0.18	0.26*
Self-explain (n = 32)		
Lexical paragraph overlap (adjacent)	0.26*	0.51**
Lexical paragraph overlap (two adjacent)	0.30**	0.59**
Noun synonym overlap	0.17	0.43**
Verb synonym overlap	0.26*	0.45**
Source evaluation (n =31)		
Lexical paragraph overlap (adjacent)	0.34**	0.51**
Lexical paragraph overlap (two adjacent)	0.37**	0.55**
Noun synonym overlap	0.30**	0.31**
Verb synonym overlap	0.11	0.27**

*$p < .05$; **$p < .01$

The results of this first set of analyses suggest that the *cohesion* of participants' constructed responses in an MD task is predictive of the quality of their subsequent integrated essay. Importantly, correlation analyses revealed that both within- and across-text cohesion indices were positively related to holistic essay quality. This suggests that the coherence-building processes enacted by students (as measured by the within- and across-text cohesion indices) were important predictors of the quality of the essays they produced in response to the MD inquiry task. However, the regression analysis indicated that across-text cohesion indices were the only significant predictors in the model; thus, the ability to produce high-quality integrated essays may be more related to the *integration* of information across MDs rather than the establishment of local connections within the documents.

Subscale essay scores and cohesion

We also investigated the relations between within- and across-text cohesion and each of the subscale essay scores (i.e., *argumentation, source use, language sophistication*, and *organization*). In general, the correlations were positive (see Table 4 for all correlations). Only one of the measures of within-document cohesion was significantly correlated with *argumentation* scores, lexical overlap (two adjacent), while all across-text indices were significantly related to *argumentation* scores. For both *source use* and *language sophistication*, only the correlations with across-text lexical overlap (adjacent and two adjacent) were reliable, indicating that participants' source use and language sophistication scores were related to the amount of across-text lexical overlap in their constructed responses. Only measures of across-text cohesion were significantly related to scores of organization, specifically lexical overlap (adjacent and two adjacent) and verb synonym overlap across-text cohesion.

We then conducted linear models to predict each of the essay subscale scores from the within- and across-text cohesion indices. In the model predicting *argumentation* subscale, scores yielded a significant model, $F(5, 87) = 6.00$, $p < .001$, $R^2 = .26$; this model had one significant predictor, across-text lexical overlap (two adjacent; $B = 0.05$, $t[5, 87] = 2.73$, $p < .01$). The model predicting *source use* produced similar results, $F(2, 90) = 5.41$, $p < .05$, $R^2 = .11$; with one significant predictor, across-text

Table 4. Correlations between lexical cohesion and essay quality subscale scores (N = 93).

Index	Within	Across
Argumentation		
Lexical overlap (adjacent)	0.16	0.38**
Lexical overlap (two adjacent)	0.20*	0.45**
Noun synonym overlap	0.08	0.21*
Verb synonym overlap	0.14	0.31**
Source use		
Lexical overlap (adjacent)	0.12	0.23*
Lexical overlap (two adjacent)	0.15	0.28**
Noun synonym overlap	0.08	0.14
Verb synonym overlap	0.06	0.18
Language sophistication		
Lexical overlap (adjacent)	0.06	0.23*
Lexical overlap (two adjacent)	0.08	0.27**
Noun synonym overlap	0.07	0.16
Verb synonym overlap	0.05	0.14
Organization		
Lexical overlap (adjacent)	0.08	0.30**
Lexical overlap (two adjacent)	0.12	0.33**
Noun synonym overlap	0.04	0.17
Verb synonym overlap	0.05	0.22*

*p < .05; **p < .01

lexical overlap (two adjacent; B = .05, $t[2, 90]$ = 2.34, p < .05). The model predicting *organization* was also significant, $F(3, 89)$ = 4.15, p < .01, R^2 = .12, but only had one predictor that was marginally significant, across-text lexical overlap (two adjacent): (B = .03, $t[3, 89]$ = 1.71, p = .09). Finally, the model predicting *language sophistication* also reached significance, $F(2, 90)$ = 4.56, p < .05, R^2 = .09; with one significant predictor, across-text lexical overlap (two adjacent) (B = .02, $t[2, 90]$ = 2.54, p = .05).

These models suggest that for all four subscale measures of essay quality, more connections made across texts were associated with better argumentation, use of sources, language sophistication, and organization in participants' essays. However, although within-text cohesion was positively related to holistic essay scores, this relationship was not evident with any of the subscales. These results provide further evidence that across-text cohesion provides greater predictive power of essay score compared to within-text cohesion. This suggests that the within- and across-text analyses were picking up the integration processes that were important for the construction of the essays across multiple dimensions.

Cohesion and experimental condition

Although the instructional condition had no significant effect on overall essay quality, these instructions likely influenced the types of comprehension processes in which participants engaged as they read (e.g., Allen et al., 2015). Thus, our second research question considered the extent to which the relations between the within- and across-text cohesion and essay scores were moderated by instructional condition. Our objective was to assess the extent to which instructions to think aloud, self-explain the texts, or evaluate the source material impacted the relations between cohesion and holistic essay quality (see Table 2 for the descriptive statistics for the cohesion indices by condition). To this end, we first computed group-level Pearson correlations between within- and across-text cohesion and holistic essay quality (Table 3).

The correlation analyses indicate that there may be differential relations between cohesion indices and holistic essay quality as a function of condition. For within-text cohesion, both the self-explain and source evaluation conditions exhibited reliable positive correlations with essay quality;

however, these correlations were not reliable in the think-aloud condition. For across-text cohesion, all conditions exhibited positive correlations with essay quality, although these were weakest for the think-aloud condition. In a follow-up analysis, we examined whether there was an interaction between the cohesion indices and condition. This interaction did not reach significance for either within-text (F[2, 87] = 2.27, p = .11) or across-text cohesion (F[2, 87] = 1.05, p = .35). Thus, while the correlational analyses reveal some interesting patterns, there was not a reliable moderation effect of condition. These inconsistent findings point to a need for future research that can more systematically determine the impact of various strategy instructions on coherence-building processes in MD contexts.

Discussion

Successful comprehension arises as a result of a reader constructing a coherent representation of the texts they have read (McNamara & Magliano, 2009). While there is considerable research on this issue in the context of single texts, how readers establish coherence in the context of understanding multiple texts warrants further research (Rouet & Britt, 2011). There remain significant challenges to assessing how coherence is achieved while individuals are reading a set of texts (Magliano et al., 2018a). The present study utilized constructed responses produced under different instructions to explore this issue. Specifically, the purpose was to examine the extent that NLP measures of within- and across-text cohesion serve as signatures of the coherence of readers' mental representations in an MD reading context.

We examined the relations between the *cohesion* of participants' constructed responses to multiple texts and the quality of their integrated essays generated after reading. We anticipated that within- and across-text cohesion of their responses would have differential effects, particularly given that the *integration* of multiple texts requires students to engage in varying forms of strategic processing. Thus, we expected across-text cohesion to relate positively to essay quality, as success on this task relies heavily on individuals' integration of information across the text set. Additionally, we examined whether these relations were moderated by instructional condition.

The results of our analyses provided evidence in support of our first prediction. Specifically, cohesion indices of participants' constructed responses were related to holistic essay quality. Consistent with the assumption that MD tasks require integration across a text set (Britt & Rouet, 2020; Rouet & Britt, 2011), both within- and across-text cohesion was positively related to holistic essay quality. These positive correlations are in line with prior work that has shown that within-text integration in think-aloud protocols is typically positively correlated with comprehension outcomes (e.g., Magliano et al., 2020). However, the current study specifically examined comprehension of information across *multiple texts*. Here, we found that across-text cohesion indices were generally more strongly correlated with both the holistic essay and subscale essay scores compared to the within-text indices. These findings suggest that participants exhibiting higher levels of across-text cohesion in their responses adopted strategies for the MD task involving integrating the texts across the entire set, rather than processing them in isolation. Some participants may have focused more of their effort on ensuring that they understood each individual text to the detriment of generating connections across the entire text set, which resulted in "over compartmentalized" essays.

Indeed, our follow-up analyses indicated that the relations between within-text cohesion were only correlated with essay scores for the argumentation subscale. This suggests that argumentation relied on participants' understanding of the individual texts, but that the other dimensions of writing quality (i.e., source use, language sophistication, and organization) more heavily relied on individuals' generation of connections across the set. Future research should attempt to identify factors that support readers in the generation of these across-text connections. For example, it may be the case that readers need to be given explicit training on strategies for how to bridge content across texts. Students with low prior knowledge of the domain may need additional scaffolds and support in integrating information across texts.

Our second research question was exploratory and examined whether instructional condition (i.e., think-aloud, self-explanation, or source evaluation) moderated the relationship between cohesion and essay quality. The correlation analyses indicated differences in the relations between cohesion and holistic essay quality across conditions. The positive relation between within-text cohesion and essay scores was only present in the self-explanation and source evaluation conditions and was not present in the think-aloud condition. The self-explanation and source evaluation instructions potentially prompted students to be more strategic in their processing of within-text connections, whereas thinking aloud may not have prompted participants to generate useful connections within the texts. This is consistent with previous work showing that prompting readers to engage in more strategic processing of a text (i.e., self-explanation) leads to higher cohesion in constructed responses compared to less strategic approaches such as paraphrasing (Allen et al., 2015).

A unique contribution of the current results is the finding that across-text cohesion was positively correlated with holistic essay scores for all three instructional conditions. These instructions may have effectively oriented participants to engage in the across-text integration necessary to write effective essays. Importantly, however, there was not a significant interaction between the cohesion indices and condition on the essay scores. We are therefore wary of placing too strong of an emphasis on the results of these condition differences. Future research should attempt to replicate this preliminary finding with a larger sample size. Additionally, research should examine whether these small instructional differences might be augmented if students were provided with substantive training on the specific strategies. In general, these results of condition differences demonstrate the importance of generating connections across documents in MD comprehension and provide evidence that these connections can be measured through the cohesion of readers' constructed responses.

The findings of the current study are important for multiple reasons. First, they provide further evidence for a theoretical link between cohesion and coherence, suggesting that cohesion cues in constructed responses can potentially provide proxies for coherence-building processes during reading. Second, our results extend prior work in this area to an MD reading context; in particular, this study attempts to more systematically examine how integration within and across texts relates to comprehension. Importantly, we found that within-document and across-text cohesion were differentially related to postreading essay scores, suggesting that integration within a single text is orthogonal to across-text integration processes. Prior work on single text comprehension has found that the cohesion of constructed responses is related to text comprehension as well as individual differences in reading skill (Allen et al., 2015). However, this work has yet to be extended to an MD context. Here, we show that constructed response cohesion can be expressed in multiple ways depending on the nature of the reading task and that these different forms of cohesion can differentially relate to learning outcomes.

These results also lend credence to the argument that MD tasks contain aspects that make them unique from single documents tasks (e.g., Rouet & Britt, 2011). Specifically, establishing within-text coherence is essential for single-document tasks that require comprehension. However, some MD tasks require readers to establish relationships across texts and with little support from the authors of the texts that make up a text set (e.g., Golding., Sousa, & van Zoonen, 2012). It may be the case that building coherent representations of individual texts in an MD context runs counter to the goals that require integration. That is, building such mental models may lead to compartmentalized representations for the individual texts. This conclusion is tentative, and the results of the present study should be replicated. However, if replicated, these results have implications and challenges for interventions that take advantage of strategies, such as self-explanation and source evaluation.

Importantly, this study is only an initial exploration of our research questions and thus has a number of limitations. Additional studies will be necessary to more thoroughly examine relations between cohesion and post-reading outcomes in MD contexts. First, we only focused on the lexical cohesion of participants' constructed responses. Research has identified several ways in which individuals can establish cohesion in a text (Crossley et al., 2019). More nuanced comparisons of different forms of cohesion (e.g., semantic, causal, or connectives) may shed light on the relations between cohesion and coherence, particularly in MD contexts. Given the

sample size of the current study, we intentionally selected a limited number of cohesion indices to reduce the Type II error rate. These indices were intended to tap into cohesion along multiple dimensions; however, they were still quite limited in scope. Future experiments with larger samples can include more measures of cohesion. Second, the current study was not sufficiently powered to explore the potential impact of individual differences, such as prior knowledge and reading skill. Thus, future studies are planned to replicate these findings and further examine whether and how instructions and individual differences moderate relations between cohesion and essay quality. Finally, further research is needed to examine the generalizability of our results across different types of text sets and different types of comprehension goals.

Overall, the current study takes an important step toward understanding the role of integration in MD comprehension tasks; in particular, it provides preliminary evidence that the cohesion of participants' constructed responses provides valuable insights into the coherence of readers' mental representations. Results of this and future work iteratively strengthen our theoretical understanding of MD comprehension processes, as well as how these processes relate to integrated essay writing performance.

Acknowledgments

Disclosure statement

No potential conflict of interest was reported by the author(s).

References

Allen, L. K., McNamara, D. S., & McCrudden, M. T. (2015). Change your mind: Investigating the effects of self-explanation in the resolution of misconceptions. In D. C. Noelle, R. Dale, A. S. Warlaumont, J. Yoshimi, T. Matlock, C. D. Jennings, & P. Maglio,(Eds.)., *Proceedings of the 37th annual meeting of the cognitive science society*.

Allen, L. K., Creer, S. D., & Poulos, M. C. (2021). Natural language processing as a technique for conducting text-based research. *Language and Linguistics Compass*, *15*(7), e12433. https://doi.org/10.1111/lnc3.12433

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

Anmarkrud, Ø., Bråten, I., & Strømsø, H. I. (2014). Multiple-documents literacy: Strategic processing, source awareness, and argumentation when reading multiple conflicting documents. *Learning and Individual Differences*, *30*, 64–76. https://doi.org/10.1016/j.lindif.2013.01.007

Barzilai, S., Tzadok, E., & Eshet-Alkalai, Y. (2015). Sourcing while reading divergent expert accounts: Pathways from views of knowing to written argumentation. *Instructional Science*, *43*(6), 737–766. https://doi.org/unh.idm.oclc.org/10.1007/s11251-015-9359-4

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). Routledge. https://doi.org/10.4324/9781315627496-6

Braasch, J. L., & Bråten, I. (2017). The Discrepancy-Induced Source Comprehension (D-ISC) model: Basic assumptions and preliminary evidence. *Educational Psychologist*, *52*(3), 167–181. https://doi.org/10.1080/00461520.2017.1323219

Braasch, J. L., Bråten, I., & McCrudden, M. T. (Eds.). (2018). *Handbook of multiple source use*. Routledge. https://doi.org/10.4324/9781315627496

Braasch, J. L. G., Bråten, I., Strømsø, H. I., Anmarkrud, Ø., & Ferguson, L. E. (2013). Promoting secondary school students' evaluation of source features of multiple documents. *Contemporary Educational Psychology*, *38*(3), 180–195. https://doi.org/10.1016/j.cedpsych.2013.03.003

Braasch, J. L. G., Rouet, J.-F., Vibert, N., & Britt, M. A. (2012). Readers' use of source information in text comprehension. *Memory & Cognition*, *40*(3), 450–465. https://doi.org/10.3758/s13421-011-0160-6

Bråten, I., & Strømsø, H. I. (2012). Knowledge acquisition: Constructing meaning from multiple information sources. In N. M. Seel(Eds.), *Encyclopedia of the sciences of learning*. Boston, MA: Springer. https://doi.org/10.1007/978-1-4419-1428-6_665

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

Britt, M. A., & Aglinskas, C. (2002). Improving students' ability to identify and use source information. *Cognition and Instruction*, *20*(4), 485–522. https://doi.org/10.1207/S1532690XCI2004_2

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). Erlbaum.

Britt, M. A., & Rouet, J.-F. (2012). Learning with multiple documents: Component skills and their acquisition. In J. R. Kirby & M. J. Lawson (Eds.), *Enhancing the quality of learning: Dispositions, instruction, and learning processes* (pp. 276–314). Cambridge University Press. https://doi.org/10.1017/CBO9781139048224.017

Britt, M. A., & Rouet, J.-F. (2020). Multiple document comprehension. In L. F. Zhang (Ed.), *The oxford research encyclopedia of education*. Oxford University Press. https://doi.org/10.1093/acrefore/9780190264093.013.867

Britt, M. A., Rouet, J.-F., & Braasch, J. L. G. (2013). 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). Routledge. https://doi.org/10.4324/9780203131268

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/10.1207/s15516709cog1803_3

Claassen, E. (2012). *Author representation in literary reading*. John Benjamins Publishing Company. https://doi.org/10.1075/lal.11

Crossley, S. A., Kyle, K., & Dascalu, M. (2019). The tool for the automatic analysis of cohesion 2.0: Integrating semantic similarity and text overlap. *Behavior Research Methods*, *51*(1), 14–27. https://doi.org/10.3758/s13428-018-1142-4

Crossley, S. A., Wan, Q., Allen, L. K., & McNamara, D. S. (2021). Source inclusion in synthesis writing: An NLP approach to understanding argumentation, sourcing, and essay quality. *Reading and Writing*. https://doi.org/10.1007/s11145-021-10221-x

Ericsson, K. A., & Simon, H. A. (1993). *Protocol analysis: Verbal reports as data*. MIT Press. https://doi.org/10.7551/mitpress/5657.001.0001

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

Gernsbacher, M. A. (1990). *Language comprehension as structure building*. Erlbaum. https://doi.org.unh.idm.oclc.org/10.2307/1423042

Golding, P., Sousa, H., & van Zoonen, L. (2012). Trust and the media. *European Journal of Communication*, *27*(1), 3–6. https://doi.org/10.1177/0267323112439366

Goldman, S. R. (2004). Cognitive aspects of constructing meaning through and across multiple texts. In N. ShuartFaris & D. Bloome (Eds.), *Uses of intertextuality in classroom and educational research* (pp. 313–347). Information Age Publications.

Goldman, S. R., Braasch, J. L. G., Wiley, J., Graesser, A. C., & Brodowinska, K. (2012). Comprehending and learning from internet sources: Processing patterns of better and poorer learners. *Reading Research Quarterly*, *47*(4), 356–381. https://doi.org/10.1002/RRQ.027

Goldman, S. R., Lawless, K., & Manning, F. (2013). Research and development of multiple source comprehension assessment. In M. A. Britt, S. R. Goldman, & J. F. Rouet (Eds.), *Reading from words to multiple texts* (pp. 180–199). Routledge. https://doi.org/10.4324/9780203131268-16

Goldman, S. R., & Rakestraw, J. S. (2000). Structural aspects of constructing meaning from text. In M. L. Kamil, P. B. Mosenthal, P. D. Pearson, & R. Barr (Eds.), *Handbook of reading research* (Vol. 3, pp. 311–335). Erlbaum. https://doi.org/10.4324/9781410605023.ch20

Graesser, A. C., & McNamara, D. S. (2012). Automated analysis of essays and open-ended verbal responses. In H. Cooper,P. M. Camic,D. L. Long,A. T. Panter,D. Rindskopf, &K. J. Sher(Eds.), *APA handbook of research methods in psychology, vol. 1. Foundations, planning, measures, and psychometrics* (pp. 307–325). American Psychological Association. https://doi.org/10.1037/13619-017

Hovland, C. I., Janis, I. L., & Kelley, H. H. (1953). *Communication and persuasion*. Yale University Press.

Karlsson, J., van den Broek, P., Helder, A., Hickendorff, M., Koornneef, A., & van Leijenhorst, L. (2018). Profiles of young readers: Evidence from thinking aloud while reading narrative and expository texts. *Learning and Individual Differences*, *67*, 105–116. https://doi.org/10.1016/j.lindif.2018.08.001

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

Kintsch, W. (1998). *Comprehension: A paradigm for cognition*. Cambridge University Press. https://doi.org/10/1998-07128-000

Landauer, T. K., McNamara, D. S., Dennis, S., & Kintsch, W. (Eds.). (2007). *Handbook of latent semantic analysis*. Erlbaum. https://doi.org/10.4324/9780203936399

List, A., Du, H., Wang, Y., & Lee, H. Y. (2019). Toward a typology of integration: Examining the documents model framework. *Contemporary Educational Psychology*, 58, 228–242. https://doi.org/10.1016/j.cedpsych.2019.03.003

List, A., & Alexander, P. A. (2019). Toward an integrated framework of multiple text use. *Educational Psychologist*, 54(1), 20–39.

Magliano, J. P., & Millis, K. K. (2003). Assessing reading skill with a think-aloud procedure and latent semantic analysis. *Cognition and Instruction*, 21(3), 251–283. https://doi.org/10.1207/S1532690XCI2103_02

Magliano, J. P., Millis, K. K., Ozuru, Y., & McNamara, D. S. (2007). A multidimensional framework to evaluate reading assessment tools. In D. S. McNamara (Ed.), *Reading comprehension strategies: Theories, interventions, and technologies* (pp. 107–136). Erlbaum. https://doi.org/10.4324/9780203810033-8

Magliano, J. P., Hastings, P., Kopp, K., Blaum, D., & Hughes, S. (2018a). Computer-based assessment of essays based on multiple documents: Evaluating the use of sources. In J. L. G. Braasch, I. Bråten, & M. T. McCrudden (Eds.), *Handbook of multiple source use* (pp. 505–526). Routledge.

Magliano, J. P., Higgs, K., Santuzzi, A., Tonks, S. M., O'Reilly, T., Sabatini, J., Feller, D. P., Kopatich, R. J. R. M., & Parker, C. (2020). Testing the inference mediation hypothesis in a post-secondary context. *Contemporary Educational Psychology*, 61(1), 101867. https://doi.org/10.1016/j.cedpsych.2020.101867

Magliano, J. P., McCrudden, M. T., Rouet, J. F., & Sabatini, J. (2018b). The modern reader: Should changes to how we read affect research and theory? In M. F. Schober, D. N. Rapp, & A. M. Britt (Eds.), *The Routledge handbook of discourse processes, second edition* (pp. 343–361). Taylor and Francis. https://doi.org/10.4324/9781315687384-18

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). Routledge. https://doi.org/10.4324/9780429439957-7

McCarthy, K. S., Yan, E. F., Allen, L. K., Sonia, A. N., Magliano, J. P., & McNamara, D. S. (2022). On the basis of source: Impacts of individual differences on multiple-document integrated reading and writing tasks. *Learning and Instruction*, 79(4), 101599. https://doi.org/10.1016/j.learninstruc.2022.101599

McNamara, D. S., Levinstein, I. B., & Boonthum, C. (2004). iSTART: Interactive strategy training for active reading and thinking. *Behavior Research Methods, Instruments and Computers*, 36(2), 222–233.

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

McNamara, D. S., O'Reilly, T. P., Best, R. M., & Ozuru, Y. (2006). Improving adolescent students' reading comprehension with ISTART. *Journal of Educational Computing Research*, 34(2), 147–171. https://doi.org/10.2190/1RU5-HDTJ-A5C8-JVWE

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). Elsevier. https://doi.org/10.1016/S0079-7421(09)51009-2

McNamara, D. S., Graesser, A. C., McCarthy, P., & Cai, Z. (2014). *Automated evaluation of text and discourse with Coh-Metrix*. Cambridge University Press. https://doi.org/10.1017/CBO9780511894664

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). Routledge. https://doi.org/10.4324/9781315687384-19

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

Millis, K. K., Magliano, J. P., & Todaro, S. (2006). Measuring discourse-level processes with verbal protocols and latent semantic analysis. *Scientific Studies of Reading*, 10(3), 251–283. https://doi.org/10.1207/s1532799xssr1003_2

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). Erlbaum. https://doi.org/10.4324/9781410603050-9

Pressley, M., & Afflerbach, P. (1995). Verbal protocols of reading: The nature of constructively responsive reading. *College Composition and Communication*, 47(2), 305–306. https://doi.org/10.2307/358808

Rouet, J.-F. (2006). *The skills of document use: From text comprehension to web-based learning*. Lawrence Erlbaum. https://doi.org/10.4324/9780203820094

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

Rouet, J.-F., Britt, M. A., Caroux, L., Nivet, C., & Le Bigot, L. (2009). The influence of story consistency and reading context on the construction of documents models from multiple sources. In *13th Biennial international conference of the european association for research on learning and instruction, EARLI*. https://hal.archives-ouvertes.fr/hal-01470833

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

Saux, G., Britt, M. A., Vibert, N., & Rouet, J.-F. (2021). Building mental models from multiple texts: How readers construct coherence from inconsistent sources. *Language and Linguistics Compass*, *15*((3), 1–19. https://doi.org.unh.idm.oclc.org/10.1111/lnc3.12409

Shermis, M. D., Burstein, J., Higgins, D., & Zechner, K. (2010). Automated essay scoring: Writing assessment and instruction. *International Encyclopedia of Education*, *4*(1), 20–26. https://doi.org/10.1016/B978-0-08-044894-7.00233-5

Snow, C. (2002). *Reading for understanding. Towards an R&D program in reading comprehension* (p. 184). RAND Corporation. https://apps.dtic.mil/docs/citations/ADA402712

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

Stadtler, M., Paul, J. M., Globoschütz, S., & Bromme, R. (2015). Watch out! - An instruction raising students' epistemic vigilance augments their sourcing activities. *Proceedings of the 37th annual conference of the cognitive science society*.

Strømsø, H. I., Bråten, I., & Britt, M. A. (2010). Reading multiple texts about climate change: The relationship between memory for sources and text comprehension. *Learning and Instruction*, *20*(3), 192–204. https://doi.org/10.1016/j.learninstruc.2009.02.001

Weston-Sementelli, J. L., Allen, L. K., & McNamara, D. S. (2016). Comprehension and writing strategy training improves performance on content-specific source-based writing tasks. *International Journal of Artificial Intelligence in Education*, *28*(1), 106–137. https://doi.org/10.1007/s40593-016-0127-7

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*, 101266. https://doi.org/10.1016/j.learninstruc.2019.101266

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

Wolfe, M. B., & Goldman, S. R. (2005). Relations between adolescents' text processing and reasoning. *Cognition and Instruction*, *23*(4), 467–502. https://doi.org/10.1207/S1532690XCI2004_2

Appendix A

Think-Aloud	Self-Explain	Source Evaluation
You will now be asked to read these texts. After you read, your understanding will be assessed with comprehension questions. **One way to help us learn how you read is to think-aloud. To help you with this task, we would like you to report your thoughts about the text while you read. Please report your thoughts that immediately come to mind regarding how you understand the meaning of the text**. After some sentences, which are bolded, there is a blank box. In the space provided below the bolded segment, please read the text, **and then write any thoughts that immediately come to mind**. Please note that there are no "right" or "wrong" **thoughts.**	You will now be asked to read these texts. After you read, your understanding will be assessed with comprehension questions. **One way to improve your comprehension is to self-explain. To help you with this task, we would like you to provide your own self-explanations of the text while you read. Please explain the meaning of the text, elaborating beyond your initial understanding of the text**. After some sentences, which are bolded, there is a blank box. In the space provided below the bolded segment, please read the text, **and then write your explanation for the meaning of the text**. Please note that there are no "right" or "wrong" **self-explanations.**	You will now be asked to read these texts. After you read, your understanding will be assessed with comprehension questions. **One way to improve your comprehension is to evaluate sources. To help you with this task, we would like you 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. After some sentences, which are bolded, there is a blank box. In the space provided below the bolded segment, please read the text, **and then write your evaluation of the source**. Please note that there are no "right" or "wrong" **evaluations.**
Here is an example from a student: **Source Information:** Roberts, Paul. *The End of Food*. New York: Houghton Mifflin Harcourt, 2008. Print. **Text segment:** Because long-distance food shipments promote fuel use and the exploitation of cheap labor, shifting back to a more locally sourced food economy is often touted as a fairly straightforward way to cut externalities, restore some measure of equity between producers and consumers, and put the food economy on a more sustainable footing. **Think-Aloud:** *"That's really interesting! I didn't know that eating locally would make things cheaper and better for the economy."*	Here is an example from a student: **Source Information:** Roberts, Paul. *The End of Food*. New York: Houghton Mifflin Harcourt, 2008. Print. **Text Segment:** Because long-distance food shipments promote fuel use and the exploitation of cheap labor, shifting back to a more locally sourced food economy is often touted as a fairly straightforward way to cut externalities, restore some measure of equity between producers and consumers, and put the food economy on a more sustainable footing. **Self-Explanation:** *"This sentence is saying eating food that is not produced locally is more expensive and that eating locally might actually reduce some costs and make the economy more sustainable. This makes sense since we spend so much money to get food shipped halfway around the world, when we could try to rely on closer food sources."*	Here is an example from a student: **Source Information:** Roberts, Paul. *The End of Food*. New York: Houghton Mifflin Harcourt, 2008. Print. **Text Segment:** Because long-distance food shipments promote fuel use and the exploitation of cheap labor, shifting back to a more locally sourced food economy is often touted as a fairly straightforward way to cut externalities, restore some measure of equity between producers and consumers, and put the food economy on a more sustainable footing. **Source Evaluation:** *"This sentence argues that eating food that comes from far away is problematic and that eating locally sourced is better, but the title of the book is The End of Food, so maybe the author is just being sarcastic or preparing an argument against eating local."*
You can see that a **think-aloud** doesn't just restate the passage. **It includes everything that comes to your mind. You can type anything that you think about the sentence as you read.**	You can see that an **explanation** doesn't just restate the passage. **It explains what the passage means. You can use anything you know about the information in the sentence to explain it.**	You can see that a **source evaluation** doesn't just restate the passage. **It considers how the source might impact what the passage means. You can use anything you know about the source and the sentence to evaluate it.**
You will now be asked to write your own **think-aloud statements** for each of the target segments within the text. Please respond to the text segments in the order in which they are presented.	You will now be asked to write your own **self-explanations** for each of the target segments within the text. Please respond to the text segments in the order in which they are presented.	You will now be asked to write your own **source evaluations** for each of the target segments within the text. Please respond to the text segments in the order in which they are presented.

Appendix B

Global Warming Text Set (Constructed responses were probed after the bolded sentences):
   Text (#1/4) Title: Source A: Health Risks for Cell Phones
   Author(s): Director Kristen Johansen
   Source: Radiation Protection Authority
   Date: 19 February 2006. The earth's climate has always changed over time. Such climate changes have until recently had natural causes such as changes in the strength of the sun, changes in the earth's orbit around the sun, and volcanic eruptions. **It now appears that for the first time mankind is facing a global climate change caused by its own activities.**

   The greenhouse effect is primarily a natural and necessary process.

   The sun has a surface temperature of approximately 6,000°C and emits various kinds of radiation. **Half of the sunrays that hit the earth's atmosphere penetrate down to the surface of the earth, the rest are reflected by clouds and other gases**. Most of the sunrays that reach the earth have short wavelengths. They warm the surface of the earth, which sends back long wavelength streams of heat. **A large proportion of these streams returned from the earth are absorbed by the clouds and the gases in the atmosphere, which then send the radiated heat back to us.**

   Some of these gases in the atmosphere are called climate gases. The most important climate gases are water vapor, carbon dioxide and methane. **They form a heat shield that slows down the radiation of heat from the earth**. This results in the surface of the earth and the air layer being heated up. **This is the same that takes place in a greenhouse where sunlight penetrates the glass panes, but radiated heat is restrained on its way out**. The result is that the greenhouse is warmer than its surroundings. Without this natural greenhouse effect the average temperature on earth would be −18°C instead of the 15°C it is today.

   In recent times, climate researchers have found that the earth's average temperature rose by approximately 0.5°C between 1850 and 2004. From around 1900 until the present day the level of carbon dioxide in the air has increased from less than 0.03% to almost 0.04%, and it appears that this increase is continuing. **This is due to the fact that we have increased our discharges of CO_2 into the atmosphere through the burning of large quantities of oil, gas, and coal.**

   Human activities have also resulted in increased discharges of other climate gases.

   This can result in more of the heat being stopped from escaping from the earth and the average temperature rising even more.

   Text (#2/4) Title: Source B: The New Cigarettes: Cell Phones?
   Author(s): Research Reporter John Friis-Gudesen
   Source: Scientific American: Magazine for Nature, Science and Engineering
   Date: 11 March 2008. The UN's climate panel concludes in its third main report from 2001 that it is highly probable that manmade discharges of climate gases have contributed significantly to the climate changes observed in the last 30 to 50 years.

   Since pre-industrial times (around 1750) the concentration of carbon dioxide (CO_2) has increased by around 31%, the concentration of methane (CH_4) has increased by around 151% and the concentration of nitrogen oxide (N_2O) has increased by around 17%. These increases are due to manmade discharges and have resulted in a stronger greenhouse effect. Human activities have also introduced into the atmosphere smaller quantities of a number of climate gases that do not exist in the atmosphere naturally.

   The increase in the concentration of CO_2 in the atmosphere forms the primary constituent (around 60%) of the strengthening of the greenhouse effect for which mankind is responsible. **These manmade discharges of CO_2 are first and foremost due to the consumption of fossil fuels (coal, oil and gas) and the deforestation of tropical regions.**

   Mankind's discharges amount to only a small part of the quantity of climate gases released into the atmosphere and the effect is minor in relation to, for example, the effect of naturally occurring water vapor. The problem is that the climate system is very complex and sensitive, and even small changes in the system can trigger major consequences. **Nature's own discharges of climate gases form part of a cycle in which, for example, rotting trees release CO_2 and living trees absorb CO_2 through photosynthesis. Our CO_2 discharges from, among other things, the burning of fossil fuels do not form part of this cycle and result in surplus CO_2 which remains in the atmosphere for a long time.**

   Text (#3/4) Title: Source C: Disastrous Side Effects from Global Warming
   Author(s): Journalist Gustav Jensen
   Source: Online Periodical: National Enquiry
   Date: 16 January 2005. **Stronger storms, more hurricanes and increasingly tumultuous weather are just a few of the negative consequences we can expect in the next few years**. Global warming may also weaken the Gulf Stream and result in serious cooling in Northern Europe.

A number of oceanographers fear highly uncomfortable side effects due to global warming. **It may weaken the ocean currents in the North Atlantic to such a degree that there is a genuine risk of serious and long-term cooling both in the Nordic Region and large parts of Europe and North America.** The Nordic Region would be significantly colder without the Gulf Stream.

Oceanographers know all too well that the warnings will cause surprise because we are reminded almost daily of the opposite, namely that global warming will raise the earth's average temperature. **However, paradoxically, both things could well occur at the same time.**

If the circulation of the Atlantic is disturbed, we could have a fall in the average temperature of 3–5°C. **This will have a dramatic effect on farming and forestry, while at the same time there will be a greater need for heating.**

And there is much that indicates that the disturbances are well underway. More ice is melting due to global warming and more precipitation is falling over Russia, among other places. **This is resulting in greater outward flows of freshwater from the major Russian rivers into the Arctic Ocean.** At the same time, we risk losing the Western Arctic ice and Greenland ice.

When the ice surrounding the poles melts, this will not just result in an increased mass of water, it will also result in increased evaporation from the oceans. This will provide hurricanes with energy. Time magazine reports that hurricanes have increased in both number and intensity since 1995.

According to the UN's climate panel, an increased greenhouse effect resulted in water levels rising between 10 and 20 cm in the last century and by 2100 ocean levels will rise by between 9 and 88 cm. This will be catastrophic for many coastal communities – especially in developing countries.

Text (#4/4) Title: Source D: Polar Exploration and Global Warming

Author(s): Journalist John Hultgren

Source: Online Periodical: Bay Area Times

Date: 13 October 2004. **Regions that are now becoming accessible due to global warming conceal enormous riches.** The melting of the ice permits the exploitation of resources in the northerly regions.

Temperatures around the North Pole are increasing at double the rate of other places around the globe according to UN experts. The Arctic ice is melting so quickly that a sea passage between the Atlantic Ocean and the Pacific Ocean may be accessible to ordinary ships during the summer by 2050. **The route through the Northwest Passage to Asia will reduce the journey distance between London and Tokyo from 21,000 to 16,000 kilometers.**

The northerly regions that are becoming accessible also conceal enormous riches. **The oil and gas deposits that are concealed there are estimated to amount to 30% of the earth's deposits.**

And there is more to be found in the northerly regions than petroleum. **There is also gold, diamonds, copper and zinc.** There will be a lot of traffic due to such exploration says Frederic Lasserre, a geographer at Laval University in Quebec in Canada who is a specialist in Arctic regions.

The director of the Nansen Environmental and Remote Sensing Center, also points out positive consequences of global warming, which occurs in the Arctic in particular: A warmer climate could result in better growing conditions and lower heating costs. The ice in the Barents Sea will be pushed northwards and eastwards due to increasing south-westerly winds and warmer weather. **This will expand winter fishing grounds and make it easier for the gas and oil industry to operate during the winter season.**

Appendix C

The rubric in the following table was used for scoring the four essay subscales

	Argumentation	Source Use and Inferencing	Language Sophistication	Organization
4	**The essay**: • discusses the sides of the argument and explicitly states a position • supports the position by providing 3+ relevant and accurate claims • supports the position and claims by providing 3+ relevant and accurate pieces of evidence	**The essay**: • explicitly refers to the majority of outside sources • synthesizes information from both within and across the referenced sources • provides deep interpretations of sources that go far beyond simple reiteration/paraphrase of the material	**The essay**: • demonstrates lexical and syntactic sophistication and variety in word choice and syntax • uses consistently appropriate word choices • demonstrates a command of English spelling, grammar, and mechanics, containing few to no errors	**The essay**: • follows a logical structure, beginning with an introduction and ending with concluding statements • is well-organized and maintains sense of flow throughout the paragraphs • is coherent and makes appropriate use of cohesive devices to signal connections between ideas

(Continued).

3 **The essay**: • discusses the sides of the argument and implicitly states a position • supports the position by providing 1–2 relevant and accurate claims • supports the position and claims by providing 1–2 relevant and accurate pieces of evidence	**The essay:** • explicitly refers to one or more of the outside sources • generates explicit connections from information within sources but generates few explicit or implicit connections across the multiple sources • provides interpretations of sources that go somewhat beyond reiteration/paraphrase of the material	**The essay:** • demonstrates lexical and syntactic sophistication but little variety in word choice and syntax • generally uses appropriate word choices • shows an understanding of English spelling, grammar, and mechanics, but may contain some errors	**The essay**: • follows a logical structure but lacks explicit introduction or concluding statements • contains evidence of organization but may lack some appropriate transitions between paragraphs • is coherent and generally makes appropriate use of cohesive devices to signal connections between ideas
2 **The essay:** • discusses the side(s) of the argument but does not provide a position • discusses the side(s) by providing 1+ relevant **or** accurate claims • supports the side(s) by providing 1+ relevant **or** accurate pieces of evidence	**The essay**: • implicitly refers to one or more of the outside sources • generates explicit or implicit connections from information within sources but fails to generate any connections across the multiple sources • relies heavily on direct quotes or paraphrases of the source material	**The essay**: • demonstrates little lexical and syntactic sophistication and little variety • contains some inappropriate word choices • shows some understanding of English spelling, grammar, and mechanics, but contains numerous errors	**The essay**: • sometimes deviates from logical structure and lacks introduction or concluding statements • contains some evidence of organization but lacks important transitions between central ideas and paragraphs • is somewhat coherent but lacks important cohesive elements
1 **The essay**: • does not discuss the sides of the argument nor provide a position • provides no claims to support a position • provides no evidence to support the side(s) or position	**The essay:** • does not refer to any of the provided sources • does not synthesize information from within or across the sources provided • fails to reference concepts described in the source material	**The essay**: • demonstrates low lexical and syntactic sophistication and little to no variety • generally does not use appropriate word choices • contains a number of spelling, grammar, and mechanics errors that render portions of the text difficult to understand	**The essay**: • generally lacks a logical sequence of thought • lacks an appropriate organizational structure • is not coherent and lacks important cohesive elements