

# Constructing Arguments From Multiple Sources: Tasks That Promote Understanding and Not Just Memory for Text

Jennifer Wiley  
Washington State University, Vancouver

James F. Voss  
University of Pittsburgh

In 2 experiments, understanding of historical subject matter was enhanced when students acted as historians and constructed their own models of an historical event. Providing students with information in a web site with multiple sources instead of a textbook chapter, and instructing them to write arguments instead of narratives, summaries, or explanations, produced the most integrated and causal essays with the most transformation from the original sources. Better performance on inference and analogy tasks provided converging evidence that students who wrote arguments from the web sources gained a better understanding than other students. A second experiment replicated the advantage of argument writing even when information was presented as an argument.

There is an important distinction that is often overlooked between remembering text and understanding it. Classroom assessment is often based on the assumption that students who remember more of a text have a deeper understanding of its content. In general, the amount of information that students recall from a text depends on the “considerateness” of the text (Beck & McKeown, 1989; Britton & Gulgoz, 1991; Mayer, 1979). In short, the easier the processing of the text, the more likely students are to remember the facts that are presented to them.

On the other hand, a number of recent studies have demonstrated that conditions that make reading more effortful can also be beneficial to learning. McNamara, Kintsch, Songer, and Kintsch (1996) found that less coherent text improved readers’ performance on inference tasks. Mannes and Kintsch (1987) found that readers performed better on a problem-solving task when an introductory outline did not match the text well. Kintsch and Young (1984) found readers were better able to use information for problem solving when they read it from more difficult expository passages than simple narrative passages. In addition, McDaniel and Donnelly (1996) found that embedding an analogy in text about a scientific principle improved understanding of the principle, but not recall of the passage. All these cases demonstrate that sometimes more effortful processing can

promote better understanding of the subject matter, although not necessarily yielding better recall of the text.

In particular, students may benefit conceptually from learning tasks that promote the construction of a situation model (Kintsch, 1994), whereas tasks that can be performed with a more superficial representation of the text, such as using a textbase, would not lead to better understanding. This distinction is consistent with the idea that the construction of mental models is the key to students’ deeper understanding of subject matter (Brown, Bransford, Ferrera, & Campione, 1983; Mayer, 1989).

It is quite likely that some kinds of writing tasks will promote mental model construction more than others. For instance, Scardamalia and Bereiter (1987) have noted a distinction between *knowledge-telling* and *knowledge-transforming* when students write essays. *Telling* is regarded as a passive transfer of information from text to paper, whereas transformation is regarded as a more active and constructive process in which the writer relates the contents of sources in new ways by making novel connections within source material as well as connections to the reader’s knowledge. Knowledge-telling thus likely involves a relatively superficial interaction with the textbase, whereas knowledge transforming may involve a more conceptual interaction with the writer’s situation model of the text contents.

In the experiments reported in this article, we investigated a variety of writing tasks to determine which specific conditions may promote the construction of situation models. Further, using inference and analogy tasks, we directly tested whether students who constructed their own models of the text could demonstrate a better conceptual understanding of the subject matter than students who engaged in less constructive activity.

## EXPERIMENT 1

Recent studies on learning from history text (Voss & Wiley, 1997; Wiley & Voss, 1996) have indicated that argument writing tasks promote more transformed, inte-

---

Jennifer Wiley, Department of Psychology, Washington State University, Vancouver; James F. Voss, Department of Psychology and Learning Research and Development Center, University of Pittsburgh.

Jennifer Wiley was supported by a postdoctoral fellowship from the National Institute of Mental Health, Grant MH 16745, at the University of Massachusetts, Amherst. We thank Anne Britt, Mara Georgi, Mike Royer, and Laurie Silfies for their help and suggestions.

Correspondence concerning this article should be addressed to Jennifer Wiley, Department of Psychology, Washington State University, Vancouver, 14204 NE Salmon Creek, Vancouver, Washington 98686. Electronic mail may be sent to wiley@vancouver.wsu.edu.

grated, and causal essays than narrative writing tasks. Further, this was especially the case when students constructed historical arguments from multiple source texts rather than from textbook-like passages. Although both students who wrote narratives from a textbook chapter and the students who wrote arguments from multiple sources performed well on a short-answer test after the essays were written, it seemed that they were performing well for different reasons. Although the students who wrote narratives had good recall of surface level information, their essays did not demonstrate a well-developed understanding of the causal structure of the historical event, nor did they make connections between the various factors that were mentioned in the source texts as well as the students who wrote arguments from multiple sources. However, in these studies there were no independent measures to test whether the students who wrote the most transformed, integrated, and causal essays gained the best conceptual understanding of the subject matter.

The purpose of the present study was to demonstrate that argument tasks not only promote the building of integrated causal models, as evidenced by analysis of essays, but also promote more conceptual understanding, as indicated by inference and analogy tasks. Because these tasks require students to go beyond the information they were given, inference and analogy tasks rely on the development of a situation model (Kintsch, Welsch, Schmalhofer, & Zimny, 1990). Thus, if the more transformed, integrated, and causal essays are related to the development of a model of the text, then students who write such essays should perform better on inference and analogy tasks, suggesting they have indeed gained a conceptual understanding of the subject matter.

In addition to testing the conceptual advantages of learning from argument writing tasks, this experiment introduced two additional writing instructions to determine what kinds of writing tasks specifically promote such conceptual learning. Two of the instructions, narrative and argument writing, were a replication of Wiley and Voss (1996). On the basis of those results, in the present study we expected to find that the writing of narratives, especially from a textbook-like source, should lead to good retention of the presented material, whereas argument writing tasks should promote the development of a causal model of the text contents and promote better conceptual understanding of the subject matter. In addition to instructions to write a narrative or an argument, students were asked to write a summary or an explanation. The summary writing condition was of interest because the instruction to write a summary should invoke synthesis of the presented material. Thus, if one advantage of writing an argument is that it requires integration in support of a thesis, then summaries might show a similar benefit on inference and analogy tasks.

On the other hand, the explanation writing condition was included because, similar to an argument instruction, it explicitly demands that writers consider why things happen. In other words, an explanation instruction may prompt a more causal account than a narrative instruction. If it is the causal and integrated nature of the argument instruction that is responsible for better understanding, then both explana-

tion and argument writing tasks should yield similar performance on inference and analogy tasks.

## Method

### *Participants*

Sixty-four undergraduates at the University of Pittsburgh participated in this experiment for credit as part of an Introductory Psychology participant pool. The sample included 41 women and 23 men. There were 36 freshmen, 16 sophomores, 7 juniors, and 5 seniors. All participants were Native English speakers. Participants were assigned randomly to condition.

### *Materials*

One half of the participants received information about Ireland from 1800 to 1850 in a web-like environment with eight separate source documents: a map, biographical accounts of King George III and Daniel O'Connell, brief descriptions of the Act of Union, the Act of Emancipation, and the Great Famine, census population data, and economic statistics between 1800 and 1850. The web-like environment used in this experiment was Sourcer's Apprentice, a program developed by Perfetti and colleagues (described in Rouet, Britt, Mason, & Perfetti, 1996) as an aid for history classrooms. Each document is represented as a book, with a title along the spine. The books are placed on a bookshelf on the main page of the program. Readers could open up to two source documents or books at the same time. There were no hypertext links between documents in this study, and readers could return to the documents whenever they wished during the writing task. The average length of each document was around 220 words. A sample document is included in the Appendix.

The other half of the participants received the same content about Ireland between 1800 and 1850, but it was presented as a textbook-like chapter (1,571 words). The information presented was identical in the two formats with the exception that the textbook format contained an introductory sentence and some (noncausal) transitional clauses at the beginning of paragraphs. Each of the eight sources constituted approximately one eighth of the textbook contents, with the passages arranged in approximately chronological order. In the web format, the sources appeared on the bookshelf in a nonsystematic order.

Following Royer, Carlo, Dufresne, and Mestre (1996), we created a sentence verification task, an inference verification task, and a principle identification (analogy) task. Each verification task consisted of 10 items, 5 true and 5 false. In the sentence verification task, students were asked to indicate whether the idea appeared in the information they read about Ireland. In the inference verification task, students were asked whether the statement was true on the basis of the information they read.

The principle identification task consisted of short descriptions of four potentially analogous historical situations that varied in surface and deep similarity to the Potato Famine: the Stock Market Crash of 1929, the Black Plague, recent Tuberculosis outbreaks, and the institution of Poll Taxes in the South after the Civil War. Students were asked to indicate on a 1-to-10 scale how similar the causes of each scenario were to what caused the significant changes in Ireland's population between 1846 and 1850.

The Stock Market crash was intended as nonsimilar in either surface or deep structure. The Black Plague, due to its natural causes and extensive loss of life, was intended as similar on a surface level, but since the plague affected all levels of society, not on a deep level. The example of recent outbreaks of Tuberculosis

among predominantly disadvantaged populations was meant to be similar on both surface and deep levels, while the institution of Poll Taxes in the South was intended as nonsimilar on a surface level, but similar at a deep level due to the socio-political inequities that allowed the famine to have such a great impact on Ireland. Two historians' ratings of the deep and surface similarity of these scenarios were consistent with these assignments.

### Procedure

Participants were given information about Ireland between 1800 and 1850 either as a textbook-like chapter or from separate sources in a web-like computer environment. Students were presented with a writing instruction page that told them, "Historians work from sources including newspaper articles, autobiographies and government documents like census reports to create histories. Your task is to take the role of historian and develop a narrative about what produced the significant changes in Ireland's population between 1846 and 1850." (for the other three writing conditions the underlined phrase was replaced with "a summary," "an explanation" or "an argument"). Students were asked to read through all the information provided before writing. (Document choosing activity was logged for the web condition. Logs indicated that all students accessed all of the documents.) For both presentation formats, students were given about 30 min to read and write and had access to the chapter or web site throughout the writing task. However, because of the continued access to the text during the writing task, we do not have independent measures for reading and writing times in these studies.

Within each presentation format, one fourth of the participants were given each writing instruction. The resulting design was a  $2 \times 4$  (Presentation Format  $\times$  Writing Instruction) with one eighth of the participants in each cell. After the writing task, participants were given the sentence verification task, inference verification task, principle identification task, a short demographic survey, and a short general history knowledge test. As no differences were found due to knowledge, this variable is not discussed further. The session lasted about an hour.

## Results

### Analysis of Students' Written Accounts

For the origin of sentences and the connective-causal terms analyses, one rater went through all of the essays and developed a coding scheme for the sentence analysis and a list of connective and causal terms. Each essay was coded according to these guidelines. A second rater coded a subset of the essays (25%) with 100% agreement. All significance levels are  $p < .05$  unless otherwise noted.

### Number of Sentences in Essays

Participants wrote an average of 12.4 sentences. No significant effects were found in the total number of sentences produced for each presentation format and writing instruction ( $F_s < 1$ ).

### Origin of Sentences in Essays

Using a classification system described in Wiley and Voss (1996), we coded each sentence for whether it was borrowed, transformed, or added information. Sentences were

coded as borrowed when they were taken directly or paraphrased from the presented material. Sentences were coded as transformed when they contained some presented information in combination with a new (not from the text) claim or fact or when they combined two or more pieces of presented information that were not connected in the text. A sentence was coded as added when it contained only novel information.

**Transformed sentences.** The proportion of transformed sentences per essay is presented for each condition in Figure 1. An analysis of variance (ANOVA) indicated significant main effects for both presentation format and writing condition, but the interaction was not significant,  $F(3, 56) = 1.44$ ,  $MSE = .05$ ,  $p < .24$ .

Students who wrote from web sources had a greater proportion of transformed sentences ( $M = .62$ ) than students who wrote from textbook chapters,  $M = .42$ ,  $F(1, 56) = 14.08$ ,  $MSE = .05$ .

The main effect for writing instruction was also significant,  $F(3, 56) = 7.18$ ,  $MSE = .05$ , with Tukey's HSD indicating that students who wrote arguments had a significantly greater proportion of transformed sentences in their essays ( $M = .71$ ) than students who wrote narratives ( $M = .39$ ) or summaries ( $M = .43$ ). The proportion of transformed sentences in the explanation essays ( $M = .54$ ) did not differ significantly from any of the other three instructions.

**Borrowed sentences.** The proportion of sentences containing borrowed information is presented for each condition in Figure 1. An ANOVA indicated significant main effects for both presentation format and writing instruction, but the interaction was marginally significant,  $F(3, 56) = 2.56$ ,  $MSE = .06$ ,  $p < .06$ .

Students who wrote from textbook chapters had a greater proportion of borrowed sentences ( $M = .41$ ) than students who wrote from web sources,  $M = .27$ ,  $F(1, 56) = 5.46$ ,  $MSE = .06$ .

The main effect for writing instruction was also signifi-

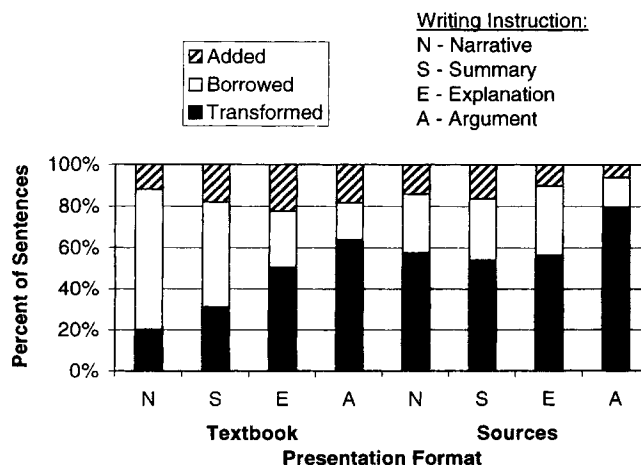


Figure 1. Origin of sentences by writing instruction and presentation format in Experiment 1.

cant,  $F(3, 56) = 4.83$ ,  $MSE = .06$ , with Tukey's HSD indicating that students who wrote narratives ( $M = .48$ ) or summaries ( $M = .40$ ) had a significantly greater proportion of borrowed sentences in their essays than students who wrote arguments ( $M = .16$ ). The proportion of borrowed sentences in the explanation essays ( $M = .31$ ) did not differ significantly from either of the other three instructions.

The marginally significant interaction seems to be due to the fact that students who wrote narratives or summaries in the textbook condition were especially likely to borrow sentences from the text,  $t(56) = 3.43$ .

**Added sentences.** No significant differences were seen in the proportion of sentences with added information (see Figure 1).

### Connections

In addition to the above analysis based on the nature of the information included in sentences, essays were also coded for the number of connective terms that were used. The connections that were coded in this analysis included inferences and causal attributions as well as temporal connections, correlations, and simple conjunctions (e.g., due to, because of, at the same time as, then, and with). Ideas that occurred in the same sentence, even without a connective term, were also coded as connected.

The means for the number of connections used in essays by each group are presented in Figure 2. An ANOVA on the number of connections in essays revealed significant main effects for both presentation format and writing instruction, but the interaction did not approach significance,  $F < 1$ .

Students who wrote from web sources ( $M = 7.4$ ) used more connectives in their essays than students writing from textbook chapters,  $M = 4.7$ ,  $F(1, 56) = 9.16$ ,  $MSE = 13.5$ .

The main effect of writing instruction was also significant,  $F(3, 56) = 3.69$ ,  $MSE = 13.5$ , with Tukey's HSD indicating

that students writing narratives ( $M = 4.4$ ) included significantly fewer connectives than students writing arguments ( $M = 8.5$ ), whereas students writing summaries ( $M = 5.3$ ) and explanations ( $M = 6.0$ ) fell between and were not significantly different than any of the other groups.

### Causal Connections

To assess the causality of each essay, an analysis was done using only causal connectives. The mean number of causal connectives for each group are presented in Figure 2. An ANOVA indicated that both main effects were significant, but the interaction did not reach significance ( $F < 1$ ).

More causal connectives were used by students writing from web sources ( $M = 3.7$ ) than by students writing from textbook chapters,  $M = 2.4$ ,  $F(1, 56) = 4.70$ ,  $MSE = 5.58$ .

The main effect for writing condition was also significant,  $F(3, 56) = 6.37$ ,  $MSE = 5.58$ . A Tukey's HSD indicated more causal connections were made when students wrote arguments ( $M = 5.0$ ) than narratives ( $M = 1.8$ ) and summaries ( $M = 2.1$ ). The number of causal connections used in writing explanations ( $M = 3.5$ ) did not differ significantly from the other three conditions.

### Summary of Writing Analyses

Although no differences were observed in essay length or the amount of added information, the other four measures revealed reliable differences between an argument writing task and a narrative writing task, and between presentation formats. Essays written from web sources had a greater proportion of transformed sentences and a lesser proportion of borrowed sentences; they also contained more connections and more causal connections in particular. Similarly, arguments contained more transformed information and less borrowed, more connections and specifically more causal connections than narratives.

The summary writing instruction yielded essays that were quite similar to narratives. Although they tended to have more connectives than narratives, summaries included significantly less transformed information and more borrowed information than arguments, and they also had significantly fewer causal connections. Thus, the summary task did not promote the building of a causal model or the transformation of the text contents in any significant way. Instead, because of the similarity of the summary and narrative results, it seems the summary instruction prompted a simple retelling of presented information rather than a more conceptual synthesis or reanalysis of the text contents.

The explanation writing task did seem to promote essays that had more transformation and causality than those of the narrative and summary writing condition, making the explanation essays closer in quality to the argument essays, but these trends did not reach significance overall. Hence, it seems that the instruction to write an explanation does not have the same effect as the instruction to write an argument.

As noted previously, learning from web sources and the argument writing instruction both tended to have similar effects on the kinds of essays that were written. It is not

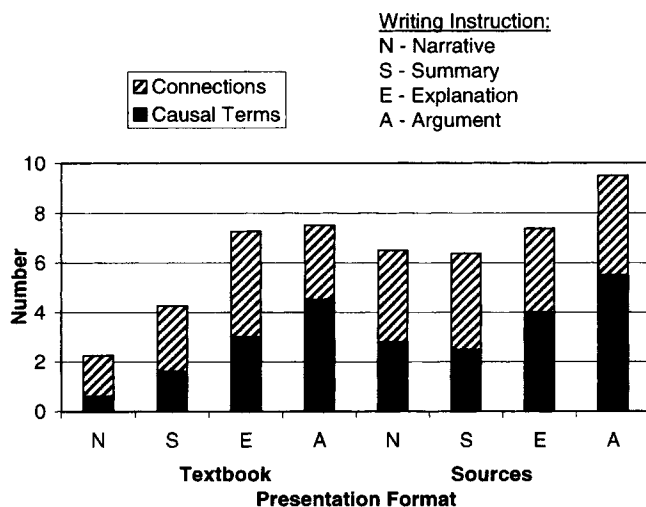


Figure 2. Number of connections and causal terms included in essays by writing instruction and presentation format in Experiment 1.

surprising then that the web source–argument condition had the highest proportion of transformed sentences, the least borrowed sentences, the most connections, and the most causal connections. Because it seemed that the argument instruction in particular gained a special advantage from the web source presentation, the effects of the writing instructions were examined in each presentation type separately. Contrasts revealed that in the web source condition, the argument instruction promoted significantly more transformations,  $t(56) = 2.62$ , marginally more connections,  $t(56) = 1.83$ ,  $p < .07$ , and more causal links,  $t(56) = 2.46$ , than the other three conditions. Meanwhile, in the textbook condition, arguments were not different from explanations on these variables, and both yielded more transformation,  $t(56) = 4.06$ , more connections,  $t(56) = 2.16$ , and causal links,  $t(56) = 3.14$ , than did narratives and summaries.

Thus, narrative and summary writing conditions promoted similar essays, regardless of whether information was presented in textbook or web format. Arguments and explanations led to similar essays when students wrote from a textbook chapter, but students in an argument writing instruction were especially likely to write more transformed, integrated, and causal essays when they wrote from web sources. As a result of the overall improvement in such measures when writing from separate sources, in the web source condition, narrative, summary, and explanation writing instructions produced similar essays, whereas the argument instruction alone produced the most transformed, integrated, and causal constructions.

*Learning Measures*

*Sentence Verification Task*

The mean accuracy scores for each condition are presented in Figure 3. An ANOVA on the accuracy scores revealed a significant main effect for presentation type,  $F(1, 56) = 3.90$ ,  $MSE = .01$ . Students reading from textbook

chapters ( $M = .75$ ) recognized presented information better than students who read from web sources,  $M = .70$ . The writing instruction main effect was marginally significant,  $F(3, 56)$ ,  $MSE = .01$ ,  $p < .07$ , as the explanation task tended to have poorer accuracy ( $M = .66$ ) than the other three conditions ( $M = .74$ ). The interaction did not reach significance,  $F = 1.46$ .

Planned comparisons based on the results of the writing analyses revealed that for students who read from a textbook, the narrative and summary instructions led to better recognition of the presented information than the explanation and argument instructions,  $t(56) = 2.39$ . No significant differences in accuracy were found because of the argument writing instruction as compared with the other three instructions in the web source condition,  $t(56) = 1.70$ ,  $p < .09$ . Thus, the narrative and summary writing tasks in the textbook condition led to particularly good recognition of the sentences that were read.

*Inference Verification Task*

The mean accuracy scores for each condition are presented in Figure 3. An ANOVA on the accuracy scores revealed a significant main effect for presentation type,  $F(1, 56) = 4.47$ ,  $MSE = .02$ . Students reading from web sources ( $M = .78$ ) recognized inferences that followed from the presented text better than students who read from textbook chapters ( $M = .71$ ). The writing instruction main effect was marginally significant,  $F(3, 56) = 2.47$ ,  $MSE = .02$ ,  $p < .07$ , as the argument instruction tended to lead to better performance ( $M = .82$ ) than the other three conditions ( $M = .72$ ). The interaction did not reach significance,  $F = 1.14$ .

Planned comparisons based on the results of the writing analyses revealed that for students who read from a textbook, there were no significant differences in performance because of narrative and summary versus explanation and argument writing instructions,  $t < 1$ . However, when students read from web sources, the argument instruction produced better performance than the other three conditions,  $t(56) = 3.12$ .

*Principle Identification Task*

The proportion of students classifying each scenario as similar to the Potato Famine is reported in Figure 4. Ratings of 1 to 5 were scored as *not similar*, whereas ratings of 6 to 10 were scored as *similar*.

The majority of students in all conditions rejected the Stock Market crash as dissimilar to the Potato Famine. Further, the majority of all groups accepted the Black Plague as similar. No differences because of presentation format or writing instruction were seen on the ratings for either scenario,  $F_s < 1.87$ .

The majority of students in all conditions tended to see the similarity of recent Tuberculosis outbreaks to the Potato Famine, but this similarity was especially noted by students in the web source condition,  $F(1, 56) = 4.00$ ,  $MSE = .25$ . There were no significant differences among the four writing instructions,  $F < 1$ .

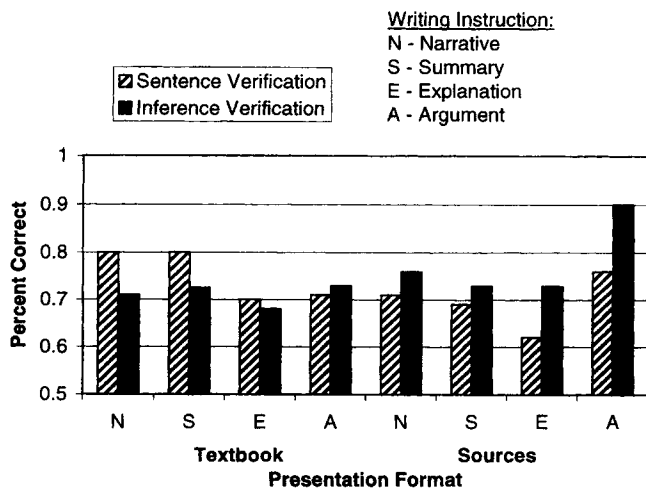


Figure 3. Performance on verification tasks by writing instruction and presentation format in Experiment 1.

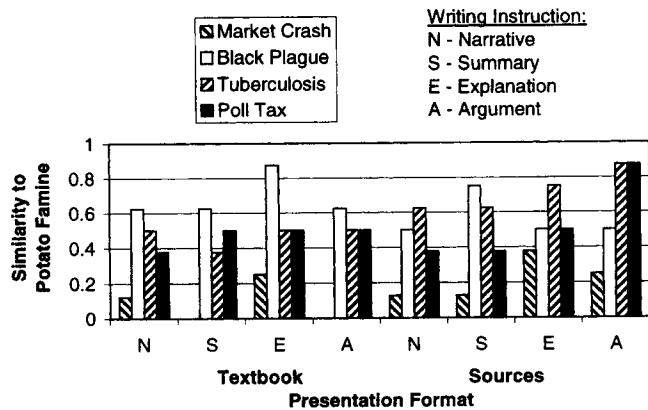


Figure 4. Percent of participants who rated scenarios as similar to Potato Famine by writing instruction and presentation format in Experiment 1.

The critical analogy was the Poll Tax scenario, which shared little surface similarity with the Potato Famine, but shared similarity in the underlying reasons for the devastating effects of the blight. Although there were no significant overall effects for presentation format or writing instruction, a planned comparison based on the writing analyses revealed that the students who wrote arguments from web sources were significantly more likely than students with the other three instructions to recognize the underlying similarity of these two scenarios,  $t(56) = 2.21$ . No differences were seen because of writing instructions in the textbook condition,  $t < 1$ . This suggests that only students in the web source/argument writing condition reliably gained an appreciation for the root causes of why the Potato Famine was so devastating to Ireland.

## Discussion

In the present study, students who wrote arguments produced essays with significantly more integration, causality, and transformation than students who wrote narratives. Most importantly, two learning measures (the inference verification task and the principle identification task) provide converging evidence that students who write arguments are gaining more conceptual and causal understanding of the subject matter than students who write narratives. Meanwhile, students who wrote narratives or summaries from textbook chapters had good recognition of sentences they had read, but they did not perform as well on the two conceptual tasks.

In addition, writing from multiple sources, in this case presented in a web-like environment, was especially beneficial to students who wrote arguments. This condition yielded the most transformed essays, as well as the best performance on inference verification and principle identification tasks. This result suggests that in order for students to gain a deeper understanding of the subject matter, writing tasks must require knowledge-transforming and not just knowledge-telling. One way to achieve this, as we have demonstrated, is to give students access to a variety of sources, and

a specific argument writing task, that requires them to construct their own take on the information they read.

We used two additional writing instructions to explore what other tasks might promote more transformation during writing, and in turn better conceptual understanding. The summary writing task, although intended to invoke more integrated accounts, produced essays similar to the narrative instruction. The explanation task, on the other hand, prompted more causal accounts than the narrative or summary instructions and did tend to produce essays similar to the argument instruction in the textbook condition but not in the web source condition. Thus, there is something special about the instruction to write an argument when reading from multiple sources, which prompts readers to construct their own representation or situation model of the text contents. Further, even though the instruction to write an explanation seems to prompt the building of a causal model, students in the web source-argument instruction are most likely to recognize the underlying causes of the Potato Famine.

One possible reason for the performance difference between the argument and explanation instructions is that arguments may be interpreted as more personal. That is, whereas an explanation may be interpreted as "tell why this happened," an argument may be interpreted as "tell why you think this happened." So, it may be that whereas an explanation could be constructed in a fairly superficial manner, an argument may require the reader to consider their own take or build their own representation of the text in a way that an explanation may not. The argument instruction may also convey that opinions need to be supported, defended, or justified. As a result, the argument instruction invites the bringing together of evidence in support of a thesis and thus the construction of an integrated causal model to a greater extent than other instructions, especially when a multiple-source presentation also promotes the building of such a model.

An alternative explanation is that it is not the causal and integrated nature of the argument writing instruction, per se, but the mismatch between the writing task and the presentation of information that contributes to the advantage seen when students write arguments. This explanation is consistent with the Mannes and Kintsch (1987) results, which indicate better conceptual learning when an advance organizer (outline) mismatched a text (Fermentation in Industry). However, there is a confound in their study, as the mismatched outline dealt with biological properties of bacteria, whereas the matched outline followed the narrative of the chapter. Because the mismatched outline is an additional source that is directly relevant to the deeper understanding that they later measure, it is not clear whether it was the introduction of additional, principled information that allowed students to perform better on the postreading sorting task or if the mismatch alone was the catalyst for developing a more principled situation model of the text.

## EXPERIMENT 2

To test whether the advantages in learning due to constructing arguments versus narratives from a single source are

dependent on a mismatch in structure between the presented information and the essay to be written, we presented the single source in the form of an argument in this experiment. A recent article from the *New York Times*, reporting on a letter written by Tony Blair to Ireland expressing regret over Britain's role in the Potato Famine, was used as a vehicle for presenting information about Ireland's history. Using the first and last paragraphs of the article as a frame, we constructed the middle section by using the exact same information about Ireland that was presented in the chapter and sources in Experiment 1. Further, in this study the single source was presented either on paper as in the previous experiment, or on a computer in a single-document web-site.

## Method

### Participants

Twenty-four undergraduates at the University of Pittsburgh participated in this experiment for credit as part of an Introductory Psychology participant pool. This included 15 women and 9 men. There were 12 freshmen, 6 sophomores, 4 juniors, and 2 seniors. All participants were Native English speakers. Participants were assigned randomly to condition.

### Materials

One half of the participants read the newspaper article containing information about Ireland between 1800 and 1850 on paper, and the other half read from a single-document web site. The information presented was identical to that which was presented in separate sources and a textbook-like chapter in Experiment 1. The newspaper version was based on an article by Sarah Lyall titled "Past as Prologue: Blair Faults Britain in Irish Potato Blight" that appeared in the *New York Times* on Tuesday, June 3, 1997.

### Procedure

Participants were given information about Ireland between 1800 and 1850 in a newspaper article, either presented on paper or in a single-document web site on a computer. Participants were presented with the same writing instructions as in the narrative and argument writing conditions in Experiment 1 and were asked to read through the article before beginning the writing task. Participants had about 30 min to read and write and had access to the article throughout the writing task.

Within both paper and computer groups, half the participants were asked to write a narrative, and half were asked to write an argument. The resulting design was a  $2 \times 2$  (Presentation Medium  $\times$  Writing Instruction) with one fourth of the participants in each cell.

After the writing task, participants were given the sentence verification task, inference verification task, and principle identification task used in Experiment 1. In addition, participants completed a short demographic survey and a general history knowledge test. No differences were found in history knowledge across conditions. The entire session lasted about an hour.

## Results

### Analysis of Students' Written Accounts

Essays were coded using methods developed in Experiment 1. A second rater coded a subset of the essays (25%) with 100% agreement.

### Number of Sentences in Essays

The average number of sentences produced was 11.1. No differences were seen either because of writing instruction or presentation format,  $F_s < 1$ .

### Origin of Sentences in Essays

Using the same classification system as in Experiment 1, we coded each sentence for whether it was borrowed, transformed, or added information.

**Transformed sentences.** The proportion of transformed sentences per essay is presented by presentation medium and writing instruction in Figure 5. An ANOVA indicated there was a significant main effect of writing instruction, while neither the presentation main effect nor the interaction was significant,  $F_s < 1$ . Students who wrote arguments had a significantly greater proportion of transformed sentences in their essays ( $M = .71$ ) than students who wrote narratives ( $M = .40$ ),  $F(1, 20) = 12.88$ ,  $MSE = .05$ .

**Borrowed sentences.** The proportion of sentences containing borrowed information is presented by presentation medium and writing instruction in Figure 5. An ANOVA indicated a significant main effect for writing instruction, whereas neither the presentation main effect nor the interaction approached significance,  $F_s < 1$ . Students who wrote narratives ( $M = .40$ ) had a significantly greater proportion of borrowed sentences in their essays than students who wrote arguments ( $M = .15$ ),  $F(1, 20) = 7.58$ ,  $MSE = .04$ .

**Added sentences.** The proportion of sentences containing added information is presented by presentation medium and writing instruction in Figure 5. An ANOVA indicated a

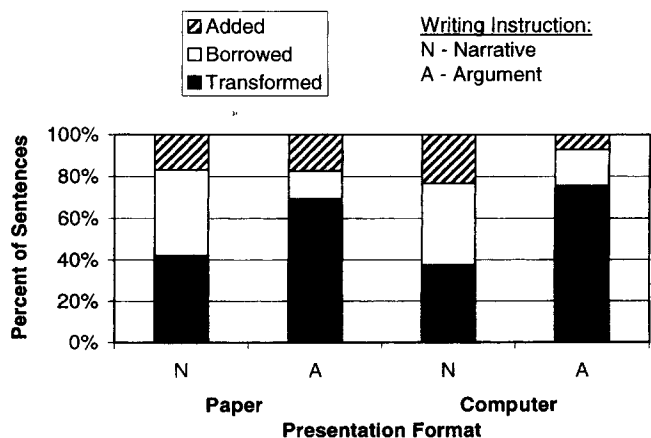


Figure 5. Origin of sentences by writing instruction and presentation format in Experiment 2.

significant main effect for writing instruction, as students who wrote narratives ( $M = .20$ ) had a significantly greater proportion of added sentences in their essays than students who wrote arguments ( $M = .12$ ),  $F(1, 20) = 5.73$ ,  $MSE = .01$ . Although the main effect for presentation medium was not significant,  $F < 1$ , the interaction between presentation medium and writing instruction was significant,  $F(1, 20) = 6.21$ ,  $MSE = .01$ . A post hoc contrast suggests that the interaction was due to the greater difference between writing instruction when students read from computers,  $t(20) = 2.50$ . Whereas the amount of added information was similar for the writing conditions when students read from paper, the narrative instruction yielded more added information, and the argument instruction led to much less added information when students read from a computer.

### Connections

The means for the number of connections used in essays by each group are presented in Figure 6. An ANOVA on the number of connections in essays revealed marginally significant main effects for both presentation medium and writing instruction, but the interaction did not approach significance,  $F < 1$ . Students writing narratives ( $M = 5.1$ ) tended to use fewer connectives than students writing arguments ( $M = 6.8$ ),  $F(1, 20) = 3.72$ ,  $MSE = 4.94$ ,  $p < .07$ . On the other hand, students who read from paper ( $M = 6.8$ ) tended to use more connectives than students who read from a computer ( $M = 5.1$ ),  $F(1, 20) = 3.72$ ,  $MSE = 4.94$ ,  $p < .07$ . This result seems due to the fact that students who read from a computer and wrote narratives used especially few connectives.

### Causal connections

A second analysis was done using just causal connectives. The mean number of causal connectives for each group are presented in Figure 6. An ANOVA indicated a significant

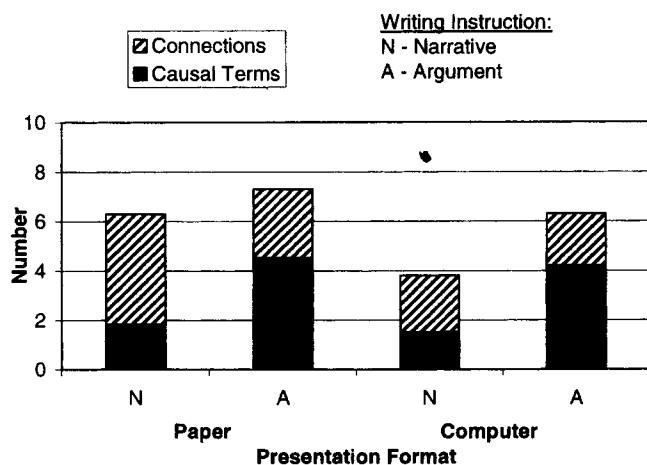


Figure 6. Number of connections and causal terms included in essays by writing instruction and presentation format in Experiment 2.

main effect for writing instruction, but neither the main effect for presentation medium nor interaction were significant ( $F_s < 1$ ).

More causal connections were made when students wrote arguments ( $M = 4.3$ ) than narratives ( $M = 1.7$ ),  $F(1, 20) = 12.80$ ,  $MSE = 3.33$ .

### Summary of Analysis of Writing

The most important result of the present study is that even with the presentation of information in the form of a single newspaper article, the writing instruction manipulation replicated the effects found in earlier studies. The argument writing instruction yielded essays with more transformed and causal information, whereas the narrative instruction yielded essays that were more borrowed and less causal.

Little difference was found in the written essays as a function of whether students read a newspaper article off of a computer or paper. Computer presentation tended to yield narratives with a bit more added information and fewer connectives. However, the proportion of transformed sentences, borrowed sentences, and causal connectives were quite similar across the two presentation mediums. Although trade-offs between paper and computer presentation of a single document may become more pronounced given a larger amount of information (cf. Haas & Hayes, 1986), for the present study it seemed to make little difference. Thus, we focus on the effects of the writing instruction in relation to the learning measures.

### Learning Measures

#### Sentence Verification Task

There were no significant differences in the recognition of sentences that had been read because of writing task,  $F(1, 20) = 1.67$ ,  $MSE = .01$ ,  $p < .21$ , or presentation format,  $F(1, 20) = 1.67$ ,  $MSE = .01$ ,  $p < .21$ , although students who read the argument from paper and wrote narratives tended to have the best recognition of the reading material,  $F(1, 20) = 3.28$ ,  $MSE = .01$ ,  $p < .09$  (see Figure 7).

#### Inference Verification Task

Consistent with Experiment 1, the argument writing task produced better identification of inferences that followed from the presented material,  $F(1, 20) = 13.7$ ,  $MSE = .01$ . As seen in Figure 7, students who wrote arguments were more accurate at the inference verification task.

#### Principle Identification Task

All participants rejected the Stock Market as similar and saw both the Black Plague and recent Tuberculosis outbreaks as similar,  $F_s < 1.80$ . However, a larger proportion of students who wrote arguments rated the Poll Tax as similar than students who wrote narratives,  $F(1, 20) = 4.63$ ,  $MSE = .23$ , as shown in Figure 8.

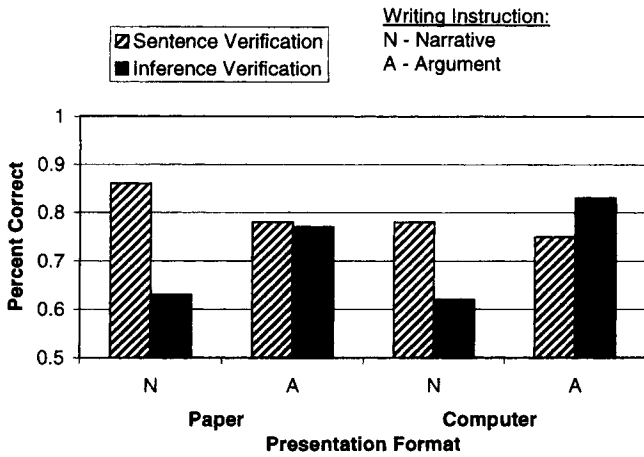


Figure 7. Performance on verification tasks by writing instruction and presentation format in Experiment 2.

Discussion

The argument writing instruction produced more causal and transformed essays as compared with the narrative instruction, even when students wrote from information presented in the form of an argument. As in previous studies, the argument instruction yielded more constructive activity as the readers wrote from presented text. In turn, students who had the most transformed essays were also more likely to identify inferences that could be made and recognize the deep causes of the historical event. The most important finding from Experiment 2 was that the advantage of writing an argument persisted even with no mismatch between the structure of the presented text and the essay that was to be written.

In addition, Experiment 2 extended the findings of Experiment 1 by suggesting another condition in which surface memory for text may be improved. In Experiment 1, sentence recognition was best when the format of the presented text most closely matched that of the essay task. In

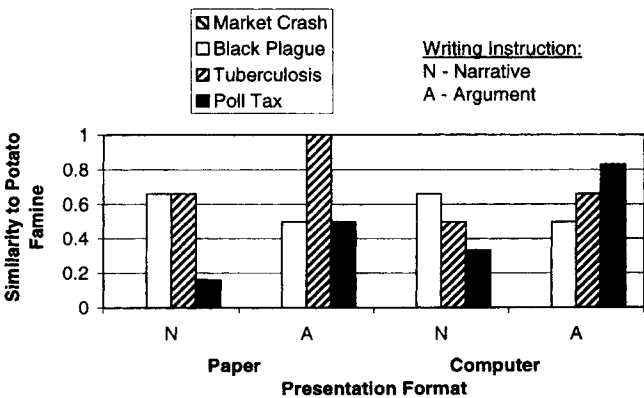


Figure 8. Percent of participants who rated scenarios as similar to Potato Famine by writing instruction and presentation format in Experiment 2.

Experiment 2, all participants read an argument about the Potato Famine, but when participants wrote narratives from a paper source, they tended to have the best surface memory.

GENERAL DISCUSSION

The present experiments are consistent with our earlier work (Voss & Wiley, 1997; Wiley & Voss, 1996) in which we found that argument writing tasks produced qualitatively different essays than narrative writing tasks in terms of transformation, integration, and causality. We suggested that students who wrote arguments seemed to gain a better understanding of the subject matter. Results of the present study support this hypothesis. In the present study, although narrative writing tasks generally yielded good retention of information, argument writing tasks were found to yield better recognition of inferences and underlying principles of the subject matter. Further, argument writing tasks were especially beneficial when students constructed historical arguments from multiple-source texts rather than from textbook-like passages.

Similar to Scardamelia and Bereiter's (1987) distinction between knowledge-telling and knowledge-transforming tasks, students seem to benefit most from contexts that require the construction of a situation model to perform the task. On the other hand, tasks that can be performed with a more superficial representation of the text, such as using a textbase, do not necessarily lead to understanding. Perfetti (1997) has suggested that multiple documents may necessitate the development of a situation model because textbase representations would be confined to single texts. The results of our studies suggest that argument writing may be another context that in particular encourages the development of a situation model. The combination of multiple sources and the argument writing task, then, makes for a learning context that highly promotes constructive activity on the part of the learner. It is this constructive activity that is correlated with better understanding.

Determining why the instruction to write an argument was such a powerful manipulation is an important direction for future research. As mentioned previously, one possible reason for the difference between the argument and other writing instructions is that arguments may be interpreted as more personal, and a greater personal investment may lead to more constructive writing. Another possibility is that the argument instruction may imply strongly that opinions need to be supported or defended. This perception may prompt students to retrieve and, most importantly, relate more pieces of information for the purposes of justification. An investigation focused on exactly how students interpret the argument instruction could be an important step in understanding why argument writing tasks yielded more constructive activity.

In addition to providing a demonstration of how to promote and measure more conceptual learning, these studies represent a much needed baseline for future work on learning from the web. The web will be an exciting place for learning research. One important implication of our work on the use of multiple sources in a web-like environment is that it is critical for teachers to give specific tasks, like argument

writing, that focus and support students' integration of information, and if possible teachers need to define a search space for their students. These studies demonstrate that even without links between sources, or differing viewpoints or conflicting information (Rouet et al., 1996), a multiple-source presentation can be seen to contribute to a more principled representation of the subject matter, especially when the student is given a task that promotes integration and transformation. It would be enlightening to know exactly what participants were doing in the multiple-source environment. With the present methodology, we were not able to analyze how long the individual sources were read, what particular parts of the texts were focused on, what parts were reexamined or compared, or what was ignored. Such an investigation of exactly which reading patterns are related to better understanding is an important next step for this line of research. Further, reader characteristics, such as reading ability, may influence the effectiveness of the learning context we studied. It may be that readers of particular ability levels benefit most from multiple-source, argument writing tasks, and this warrants further investigation.

More generally, the present studies underscore that conditions that promote retention of information are not necessarily the same contexts that promote better understanding. Teachers should be aware that multiple-choice tests and short-answer quizzes that emphasize recall of facts without requiring deeper thought may actually bias against students who engage in deeper processing. Especially when we ask students to engage in transforming tasks, we must be sure to measure their learning with appropriate measures that encourage them to think deeply and construct their own models of the texts they read.

### References

- Beck, I., & McKeown, M. (1989). Expository text for young readers: The issue of coherence. In L. B. Resnick (Ed.), *Knowing, learning, and instruction: Essays in honor of Robert Glaser* (pp. 47–65). Hillsdale, NJ: Erlbaum.
- Britton, B., & Gulgoz, S. (1991). Using Kintsch's computational model to improve instructional text: Effects of repairing inference calls on recall and cognitive structures. *Journal of Educational Psychology, 83*(3), 329–345.
- Brown, A., Bransford, J., Ferrara, R., & Campione, J. (1983). Learning, remembering and understanding. In J. Flavell & E. Markman (Eds.), *Handbook of child psychology: Vol. 3. Cognitive development* (4th ed., pp. 77–166). New York: Wiley & Sons.
- Haas, C., & Hayes, J. R. (1986). What did I just say? Reading problems in writing with the machine. *Research in the Teaching of English, 20*(1), 20–35.
- Kintsch, W. (1994). Text comprehension, memory, and learning. *American Psychologist, 49*(4), 294–303.
- Kintsch, W., Welsch, D., Schmalhofer, F., & Zimny, S. (1990). Sentence memory: A theoretical analysis. *Journal of Memory and Language, 29*, 133–159.
- Kintsch, W., & Young, S. (1984). Selective recall of decision-relevant information from texts. *Memory and Cognition, 12*(2), 112–117.
- Mannes, S., & Kintsch, W. (1987). Knowledge organization and text organization. *Cognition and Instruction, 4*, 91–115.
- Mayer, R. (1979). Twenty years of research on advance organizers: Assimilation theory is still the best predictor of results. *Instructional Science, 8*(2), 133–167.
- Mayer, R. (1989). Models for understanding. *Review of Educational Research, 59*(1), 43–64.
- McDaniel, M., & Donnelly, C. (1996). Learning with analogy and elaborative interrogation. *Journal of Educational Psychology, 88*, 508–519.
- McNamara, D., Kintsch, E., Songer, N., & Kintsch, W. (1996). Are good texts always better? *Cognition and Instruction, 14*, 1–43.
- Perfetti, C. (1997). Sentences, individual differences, and multiple texts: Three issues in text comprehension. *Discourse Processes, 23*, 337–355.
- Rouet, J. F., Britt, M. A., Mason, R., & Perfetti, C. (1996). Using multiple sources of evidence to reason about history. *Journal of Educational Psychology, 88*, 478–489.
- Royer, M., Carlo, M., Dufresne, R., & Mestre, J. (1996). The assessment of levels of domain expertise while reading. *Cognition and Instruction, 14*(3), 373–408.
- Scardamalia, M., & Bereiter, C. (1987). Knowledge telling and knowledge transforming in written composition. In S. Rosenberg (Ed.), *Advances in applied psycholinguistics: Vol. 2, Reading, writing and language learning*. Cambridge, England: Cambridge University Press.
- Voss, J. F., & Wiley, J. (1997). Developing understanding while writing essays in history. *International Journal of Educational Research, 27*, 255–265.
- Wiley, J., & Voss, J. (1996). The effects of "playing" historian on learning in history. *Applied Cognitive Psychology, 10*, 63–72.

## Appendix

## Sample Document

**The Great Famine**

The Great Famine had its origins on the eastern seaboard of the United States and Canada, where potato crops were ravaged by blight in 1842. The blight was a fungus infection that made crops unsalvageable. In August 1845 the disease made its first European appearance in the south of England, and by September it had crossed to Waterford and Wexford from which it spread to about half of England. By the time it reached Ireland nearly one-sixth of the potato crop had already been dug, and thus escaped damage.

It was not until 1846 that the blight made a serious impact. The crop of that year was almost a total loss. The blight abated in 1847, but many people had been forced to eat their "seed" potatoes to survive the previous winter. The yield in 1847 was only 2 million tons. The year 1848 saw the number of acres sown nearly triple, but

a cold, wet growing season yielded a poor crop with only half the yield per acre of the previous year. Real recovery began only in 1849. There was a small decrease in acreage sown, but the yield was over 4 million tons.

The blight was especially devastating to Ireland because, for the majority of the Irish population, potatoes were the mainstay of their diet. In general, only 20% of the potato crop was grown for export. In contrast, the vast majority of other crops, such as wheat or oats, were grown for export. These were considered "rent" crops and were sold in order to pay rent on the land.

Received January 5, 1998

Revision received June 15, 1998

Accepted August 28, 1998 ■