

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

CHAPTER 7

DEVELOPING UNDERSTANDING WHILE WRITING ESSAYS  
IN HISTORY

JAMES F. VOSS and JENNIFER WILEY

Learning Research and Development Center, University of Pittsburgh, Pittsburgh, PA 15260,  
U.S.A.

Abstract

Using the textbase-situation model of discourse processing and assuming a distinction of learning (recall of text contents) and understanding (relating different parts of text contents or text to non-text contents), it was found that individuals reading text contents from a number of sources who wrote an argumentative essay about the contents and then rated content elements for importance developed a better understanding of the contents than individuals writing a narrative essay and making importance ratings either before or after writing. © Elsevier Science 1997

The decades of the 1980s and 1990s have been marked by the study of learning and reasoning in various subject matter domains (Voss, Wiley, & Carretero, 1995). While such domains include non-school subjects such as chess, much of the research has been focussed upon the subject matter of the school curriculum, especially physics and mathematics. More recently, however, the topic of history has received increasing attention (Carretero & Voss, 1994), this chapter summarizes the results of a study conducted on this topic.

The study of subject matter learning and reasoning has in part been motivated by the desire to develop a greater understanding of how people learn, reason, and think. But the motivation has also been pedagogical: there is the desire to improve classroom instruction by learning more about how students think in subject matter terms. This motive is quite strong in the United States because national and international studies suggest American students are not acquiring appropriate knowledge and skills. In the field of history, published reports (e.g., Beatty, Reese, Perisky, & Carr, 1996; Ravitch & Finn, 1987) suggest student knowledge of history generally tends to be poor.

Cognitive Framework

It is assumed that learning is positively related to the level of processing that occurs when an individual relates new input to his or her pre-existing knowledge. Thus, because experts are better able to integrate new information with well-developed knowledge of the subject matter, experts

tend to retrieve information better than novices. The Kintsch (1994) conceptual-integration model captures this idea by postulating the existence of a situation model, which involves the activation of information in memory by the input and the integration of input contents with the activated contents of memory.

Often, however, as when reading a novel, a person may have the general characteristics of a model in memory and some knowledge of the time and location at which the story takes place, but the person, during the course of reading the novel, needs to construct a model of the plot and the characters of the novel, modifying the model as the reading continues. To include in the model all of the information in the novel is futile, in part because of limitations of working memory. As a result, the individual usually selects information in order to maintain coherence, establish causality, maintain impressions of the characters, and perform other functions.

How, then, is this process different from the process used when a student reads a history text? Based upon the results reported in this chapter and those of another study reported elsewhere (Wiley & Voss, 1996), two differences can be noted. One is that history text can occur in different forms and processing may vary with the nature of the presentation. Specifically, the material may be presented in a textbook, in a volume on a given historical topic, or in a history journal. Historical information also can be presented via records, newspaper articles and editorials, paintings, photos, and other artifacts (collectively known as sources). A question of interest then is how the nature of the source influences the processing that occurs. In this case the "history" needs to be constructed from the historical information, with such information requiring selection and integration.

A second way in which the study of history is different from reading a novel is that, in the school context, the reading of history is typically followed by some type of questioning of the student about the contents of the history assignment. The task can take the form of a multiple-choice test, the writing of an essay about a topic discussed in the assignment, the defending of a particular position about the interpretation of the historical content, or some other procedure involving performance assessment.

The above considerations suggest a distinction between the concept of learning and that of understanding. For our purposes learning will refer to a person's ability to perform on a task that measures the acquisition of some content. Thus, what and how much one remembers from the contents of a history chapter would define what the person has learned. However, understanding is taken to refer to the knowledge a person has about the underlying conceptual relations of a given topic, the relations often including the interpretation of the presented material. A person therefore may learn quite a bit in terms of recall, but have a poor understanding, in terms of understanding important conceptual relationships related to the material. Moreover, it is assumed that learning can take place with a relatively low amount of processing whereas understanding generally involves more extensive processing, with the latter also involving greater integration of new and old information.

Wiley and Voss (1996) conducted a study in which essentially the same information was presented in two different formats. Students were asked to read material about the Irish potato famine of the mid 19th century, with the material presented either in a standard textbook format or as sources. The material included charts and graphs of population size and immigration figures, as well as social, political, and historical information. It was hypothesized that learning from sources would yield higher performance than learning from a textbook because presumably more processing would be required to integrate the source material than to integrate the material in the textbook since the textbook information was already organized. The source information, however, required conceptual integration and this required processing.

A second variable manipulated was the assigned task. The students were asked to write an essay, with each of three groups writing a different type of essay. All three groups were asked to indicate in their essay what produced the significant changes in Ireland's population between 1845-1850, with one group being asked to write a narrative, another being asked to write a history, and one group being asked to write an argument of why this occurred. The hypothesis was that writing an argument would require the most processing because that task required examining possible factors contributing to the population changes and organizing them into a reasonable argument. Writing a narrative was expected to involve less processing, while writing a history would depend upon the writer's idea of what a history is.

While the test hypotheses pertained to the main effects of the two variables under study, the primary focus was upon their interaction. It was hypothesized that the deepest or most extensive processing would occur in the condition in which students read the sources and wrote an argument. The effects of the two variables would sum, thereby producing the most extensive processing. Writing a narrative essay after reading the materials in textbook-like form was also expected to fit well together since the material and essay organizations were highly similar. However, because they mapped so well onto each other, processing would not be extensive. Thus, it was hypothesized that while this condition would produce substantial learning in terms of recall of presented information, the processing would result in limited understanding. In other words, while both textbook-like presentation of material with narrative essay condition and the source presentation with argument writing would both produce good recall of information, only the latter condition would produce high levels of understanding. Again, the results involving the history essay condition was expected to depend upon the students' concept of a history text.

The results generally supported the hypotheses. While the textbook presentation/narrative essay condition and the source presentation/argument essay condition produced better recall than the other four conditions and recall did not vary between the two conditions, the source presentation/argument essay condition yielded superior performance on measures of understanding such as the number of connections made between textual factors and the number of causal links stated in the essay.

#### Present Experiment

The experiment reported here is concerned with how individuals select information from sources and use it to write an essay. The procedure used in the present study was to ask the students to read information about the Potato Famine and write an essay of a particular type. When writing the essay the students were allowed to view all presented information, primarily because we wanted the focus of the work to be on essay writing, not memory. Because the students did not know the essay condition to which they were to be assigned as they read the presented information, there should be no differences during initial reading in relation to the essay condition manipulation. Given the results of Wiley and Voss (1996), it was hypothesized that individuals in the argument essay condition would develop a causal model of the Irish population changes as they defend their position, with their essays including more connections of concepts and more causal connections than essays in the narrative and history conditions.

In addition to the essay task, students were asked to rate the importance of each statement of the presented text, with students rating the 70 statements either before or after writing the essay. This manipulation was carried out to test the hypothesis that students would write better essay if they first indicated the importance of the specific items of textual information.

It further hypothesized that the effect of rating the importance of items on essay quality would vary with the essay condition. In the argument condition, having students first indicate what is important would disrupt the development of the causal model by constraining the development of the model when it is being organized and written. Understanding should then be reduced in this condition. In the narrative and history essay conditions, however, such disruption would not occur because a causal model is not being developed. Moreover, it could be hypothesized that in the history and narrative essay conditions what students do may be to select the important contents and write about them, essentially the same thing that they may do when they are given an importance rating task. Finally, it was hypothesized that items students used in their essays would receive higher ratings than those not used, and by comparing the ratings given before and after essay writing it would be possible to determine whether the writing of the essay produced changes.

### Method

Ninety-six undergraduates at the University of Pittsburgh participated in this experiment for credit as part of an Introductory Psychology subject pool. All participants received information about Ireland from 1800 to 1850 in the form of eight separate sources, including a map, biographical accounts of King George III and Daniel O'Connell, brief descriptions of the Act of Union 1801, Act of Emancipation 1829, and the Great Famine; census data on the population size, the death rate, and the emigration rate between 1800 and 1850; and economic statistics on crop selling prices, rent costs, distribution of land holdings, and occupational breakdowns between 1800 and 1850.

An importance rating task was created by extracting the basic ideas from each of the sources (excluding the map) yielding 70 basic idea units. The idea units were listed on a page in random order. At the top of the page, students were asked "How important were the following points toward producing the significant changes in Ireland's population between 1846 and 1850?" They were then presented with a ten-point scale in which "1" was defined as "Not at all Important" and "10" as "Extremely Important." A short-answer 20-item general knowledge test was also given, the test containing questions such as "What did Gutenberg invent around 1450?" and "In what country did the Boxer Rebellion of 1900 occur?" Participants were given packets containing the separate sources about Ireland from 1800 to 1850. After reading through the information, one-half of the students in each essay condition were presented with the importance rating task and then a writing task. The other half performed the writing task before the importance rating task. The writing task had the following instructions: "Historians work from sources including newspaper articles, autobiographies, and government documents like census reports to create histories. In this packet there are a number of documents about Ireland between 1800 and 1850. Your task is to take the role of historian and develop a *history* about what produced the significant changes in Ireland's population between 1846 and 1850. You will have about 30 minutes for this task. You are expected to make full use of that time."

One-third of the students saw the above instructions. For another third, the underlined word was replaced with *narrative* and for the remaining third the underlined word was replaced with *argument*. The resulting design is a 2x3 (task order x writing instruction) with 16 participants in each cell.

After completing both the importance rating and writing tasks, students completed a short questionnaire that requested information such as age, sex, educational status, number of college

history courses taken, and amount of interest in history. They then completed a general history knowledge test. The students were in groups and each session lasted about one hour.

### Results

Understanding of the presented material was assessed through analyses of the structure and content of the written accounts. As no differences were found in history knowledge across either the type of essay or the locus of the importance rating conditions, ( $F_s < 1$ ), this variable is not discussed further.

### Analysis of Students' Written Accounts

Three general aspects of student's writing were considered: the structure of their accounts, the integration of the information to be included in the accounts, and the selection of that information. Overall, analyses of the students' accounts indicated that both the essay type and timing of the importance ratings had an effect on the way students organized their written accounts and transformed and integrated the content within their accounts, especially with regard to causal relations. Specific analyses of students' writing included (a) the organization or structure of the account (i.e., collection of ideas versus causal essay), (b) the connection of idea units (i.e., the extent to which students recognized the possible relations between the ideas that were presented), (c) amount of explanation (i.e., number of causal connectives), (d) the origin of the information contained in sentences (i.e., information taken directly or paraphrased from sources, versus transformed or completely novel information), (e) the exhaustivity of the account (i.e., the extent to which the idea units mentioned in the text were included in students' accounts), and (f) importance of information included (i.e., the extent to which the idea units included in the essays were rated as the most important).

### General Description of the Essays

The 30 minute writing task produced essays with an average of approximately 14 sentences. There were no differences in the length of essays due to writing task,  $F=1.19$ , but there was a significant difference in length depending on whether the essays were written before or after the importance rating task. Students wrote longer essays when essays were written before the importance ratings were made,  $M=15.33$ , than when written after the importance ratings were made,  $M=11.92$ ,  $F(1,90)=9.76$ ,  $p<.002$ . This result suggests that judging the importance of particular components of the passage acts to constrain the amount of information incorporated into the essay.

### Measure of Essay Structure

Using Meyer's (Meyer, 1985) taxonomy, all essays were classified as either having a collective structure (that is, the essay consisted of a loosely or temporally associated listing of ideas)

taken directly from the presented text, "added" if the contents were not in the presented text, "transformed" if the sentence contains a combination of "borrowed" and "added" information or if it combines information presented in the text in a new way. The proportion of transformed sentences was .50 for the argumentative essay group, .33 for the narrative group and .40 for the history group,  $F(2,90)=7.27$ ,  $p<.001$ . A Tukey's test indicated that the argument essays contained significantly greater proportion of transformed sentences than the narrative essays.

### Exhaustivity

Exhaustivity was measured by the number of idea units out of the 70 in the presented material that were included in the essay. More items were used in essays when the essays were written before the importance ratings,  $M=17.71(6.08)$ , than when the essays were written after the importance ratings,  $M=14.94(5.21)$ ,  $F(1,94)=6.01$ ,  $p<.02$ . Essay type was not significant,  $F(2,93)=1.05$ , while the interaction of the two variables was significant at the .09 level,  $F(2,90)=2.43$ . This near significance is reflected primarily in the argumentative essay condition, which the number of idea units included when the essay was written before the importance ratings was 19.62, but when the essay was written after, the mean was 13.31. The corresponding means for the narrative and history conditions were 18.19 and 16.50 (before) and 15.50 and 15.00 (after). This result is consistent with the idea that the rating procedure tended to inhibit the development of argumentative essays.

### Importance Ratings

The mean overall importance ratings did not vary with respect to when importance ratings were made,  $F<1$ , essay type,  $F<1.28$ , or the interaction of the two variables,  $F<1$ . However, those items of the text that were used by a given writer in an essay were rated as more important than those not used in the essay  $F(1,90)=165.31$ ,  $p<.0001$ , with the mean of used items 7.76 and unused items 5.21. While there was little difference in importance ratings of unused items related to when the ratings were made, used items were rated as significantly more important when ratings were done after essays were written,  $M=8.15$ , than before,  $M=7.37$ ,  $F(1,90)=4.73$ ,  $p<.03$ . This indicates that using items in an essay increased the perceived importance of those items. While essay type and the interaction between essay type and the timing of essay were not significant,  $F_3<1$ , there was a significant interaction between whether an item was used or not used and essay type,  $F(2,90)=5.26$ ,  $p<.007$ . The means for the three essay types for the not-used items were narrative 5.45, argumentative 5.29 and history 4.91, while for the used items the respective means were 7.13, 8.03, and 8.13. Thus, the narrative essay condition had the highest proportion of unused items and the lowest rating for the used items, suggesting that individuals in the narrative condition tended to include more less-important statements in the writing of their essays, while the importance rating data, therefore, supported the hypotheses that in writing essays individuals intended to use the items they considered to be most important, and when an item is used it is perceived as more important than it is before it is used.

### Discussion

The present study replicated some of the findings of an earlier study (Wiley & Voss, 1996), namely, that individuals who write arguments from sources (without having made importance ratings) produced essays that included more causal connections and transformations of presented material than individuals who wrote a narrative or a history. Essays written in the argument condition were also more likely to have a causal instead of a collective, list-like structure than essays written in the narrative and history conditions. These results, again, suggest that individuals in the argument condition are constructing causal models to a greater extent than the individuals in the other two conditions.

Further, the present study demonstrates that an importance rating task can disrupt the building of causal models in the argument condition. When the importance rating task preceded the essay task, the argument condition no longer had significantly more connections, causal connections, or transformations than the other two conditions. Nor were there more causal structures than collective ones in the argument condition when the rating task preceded essay writing. Thus, rating the importance of specific items may constrain the integration and organization of text information, and the relating of it to information in memory. It also reduced the number of concepts employed overall in essay writing.

An interesting twist is that although the importance rating task seemed to disrupt the integration of information in the argument condition, no decreases were seen in the other two conditions. In fact, there is some evidence that the importance rating task may have facilitated the integration of source information in the narrative and history conditions, as the number of causal connectives was greater in these conditions when importance ratings were made first than when they were after writing the essays.

The present results relate to the distinction that has been made between "knowledge-telling" and "knowledge-transforming" writing processes (Bereiter & Scardamalia, 1987) or more generally to the amount of constructive activity that occurs during writing. Bereiter and Scardamalia have described the writing process of more and less skilled writers, finding that while less skilled writers often rely on the explicit assigned topic or genre to structure their writing, more skilled writers are more likely to have a main point or structure "emerge" from their thinking about the textual information (Bereiter, Burtis, & Scardamalia, 1988). And, whereas less-skilled writers use textual information to fill in the "slots" of a generic discourse schema, more skilled writers create novel frameworks based on consideration of both the specific content and task, and then adapt or "transform" the textual information to cohere within those frameworks. Hence, the less skilled writers are basically reporting what they have been told, perhaps with some surface re-arrangement, in a process involving little constructive activity that has been seen as "knowledge-telling." On the other hand, more skilled writers select, re-organize, integrate, and synthesize textual information, in a process involving a great deal of constructive activity that has been called "knowledge-transforming."

One way of viewing the present study is as an attempt to find conditions under which students are more likely to engage in "knowledge-transforming" or constructive activity while writing. Our results suggest that the argumentative essay task provided for more knowledge transformation than the other two essay instructions. This seems to be because developing an argument involves selecting and organizing information as well as transforming it and relating it to information in memory. This process can be quite a constructive and creative act. While writing a narrative also may be, students in the Wiley and Voss (1996) study writing from a text already had the narrative present, while those writing from sources in both studies apparently emphasized the