

A Case Study of Developing Historical Understanding via Instruction

The Importance of Integrating Text Components and Constructing Arguments

James F. Voss and Jennifer Wiley

The purpose of this chapter is to review a set of studies concerned with enhancing student learning and understanding of history.¹ The studies have focused primarily on two questions: whether presenting segments of a history text separately (as multiple texts) produces better performance than presenting the segments as a single text and whether writing an argumentative essay about a historical topic produces better performance than writing other types of essays, such as a narrative. The interaction of these two factors was also of interest. Before providing the rationale for the research, however, we need to consider the concepts of learning and understanding as we are using them.

Learning and Understanding

Learning

From a psychological perspective, the concept of learning is most frequently the acquisition of particular information or the development of some type of skill by an individual, although the term may also be used in relation to the acquisition or change of beliefs, attitudes, or other concepts. In the context of history, the acquisition might consist of information about the American Civil War, the development of skill in writing historical arguments, or a change in one's attitude about the

policies of the United States toward American Indians during the settling of the West.

Relatively standard types of evaluation are employed in the classroom to determine whether a student has learned the information under study. These include multiple-choice, completion, and short answer tests, as well as exams requiring the student to write longer essays. Students may also be required to write papers on topics related to the subject matter under study. For the purpose of the research reported in this chapter, the term "learning" will refer to the ability to recall or identify correctly the contents of a text. Specifically, we measured learning by determining the number of correct answers given by participants to questions about the text contents and by asking participants to verify whether particular statements were true according to the text contents.

Understanding

There is less precision and agreement in the use of the term "understanding" than there is for the term "learning." Most psychologists would probably agree that the term "understanding" implies not only knowledge of a given object, issue, event, or person but also knowledge of components, causes, or underlying operations that pertain to the issue in question. There can be differences in what two or more people understand about the same issue, yet all may feel that they "understand" that issue. This arbitrariness correctly suggests that what constitutes understanding, or how well something is understood, is open to debate.

At an intuitive level, if a person says she understands how a car runs, we think she knows the parts of the car and how they are involved in the car's operation. If a person says he understands another person, we think he knows the person reasonably well and is able to tell what is producing a particular aspect of the other person's behavior. If a person states that she understands the American Civil War, we think she means that she has a substantial knowledge of the war, what produced it, how and why it was fought as it was, and why it concluded as it did, along with perhaps knowledge of the war's consequences and an awareness of any controversies that exist with respect to any of these topics. However, in the case of history, no matter how the person describes such an understanding, someone else may state that the war is not really properly understood unless one considers some other factor(s) that have not been considered or not been considered appropriately.

In order to describe how we measured understanding in our research, we first need to consider the theoretical basis of the work. When an individual reads a text, the person develops what Kintsch has termed a situation model.² This model is a mental representation of the text contents that takes into account what the individual already knows about the topic of the text. Thus, if a person reads that in the year 2000 the Democrats will nominate Senator Dianne Feinstein of California for vice president, the person's mental representation may include not only this statement but the fact that California has more electoral votes than any other state and that Feinstein's nomination may help the Democrats to carry the state in the November 2000 election.

Two points may be made with respect to a situation model and a person's knowledge. As would be expected, people with more knowledge about a particular topic typically produce a more developed representation. Also, knowledge utilization is important. It is not sufficient just to have the knowledge; it must also be accessed and used.

At a more molecular level, the situation model may be considered as a node-link network structure, with the nodes representing the concepts and the links denoting the relations between concepts. Moreover, there are different types of links, that is, different types of relations. For example, there may be a link indicating that Stonewall Jackson was a general in the Confederate army. This link also may point to a category membership, indicating that there were a number of Confederate generals. Another type of link is causal; these links indicate a causal connection, such as "Lee lost the battle because he overestimated the strength of the Union troops." A number of expressions in English portray some level of causal relationship. These terms include "influences," "produces," "leads to," "causes," "is due to," and "is attributable to." Such expressions not only cue the reader to the relation; they also provide a means of obtaining an approximation of the structure of the overall text. Viewed in this context, understanding is assumed to refer to the type and extent of the node-link structure or, correspondingly, to the nature of the contents of the mental representation that a person has and utilizes in a given context. Having a better understanding in this context requires having a good sense of the causal relations and other relationships that affect the topic in question.

Thus far, we have considered the idea of a mental representation or a situation model and its underlying structure. We now turn to the issue of process. We assume that when individuals read a text or perform any of a number of other tasks, they may process the incoming information either

at a relatively shallow or superficial level or at a deeper level. In terms of the effects of such processing differences, superficial processing leads to a less developed situation model; it is as if much of the information a person has in memory about the topic in question is not activated when the text is read, and the memory information therefore is not linked to the input, nor does it become part of the mental representation of the incoming information. On the other hand, deeper processing presumably activates more related information, and a more extensive mental representation is therefore developed. The deeper processing should then lead to a representation with more causal relations and connectives. An important point regarding superficial and deeper processing is that more superficial processing should lead to a substantial knowledge of the surface structure of the text, but it should not lead to a relatively substantial understanding. On the other hand, deeper processing should produce both a substantial knowledge of the text structure and a more developed understanding.

Given these considerations, we now consider the five measures of understanding employed in the present research. Three types of measures were obtained via analyses of the written essay. The essays were first analyzed for their overall organization. Using a method of analysis developed by Meyer,³ each essay was categorized according to whether it primarily listed information with minimal focus or was more analytic, having a thesis or conclusion, and was organized in relation to that conclusion. The second occurrence was taken as a measure of better understanding.

The second essay measure involved connectives. The number of times connective words were used in each essay was taken as a measure of understanding; the connectives included inferences, temporal links, conjunctions, and causal links. In addition, the number of causal connectives alone was determined for each essay. A greater number of causal connectives was assumed to indicate better understanding. Since the number of all connectives was in substantial agreement with the number of only causal connectives, we will discuss only the results obtained for the causal connectives.

A third essay-related measure was sentence origin. Using a procedure employed by Greene,⁴ we placed each sentence of each essay in one of three categories: borrowed, added, or transformed. Borrowed sentences were those taken directly from the text. Added sentences were statements that had no text content. Transformed sentences either had content that included both text content and additional information the reader

brought into the text situation or brought together two units of the text that had been previously unconnected. Transformed sentences were assumed to depict better understanding because they showed a rephrasing of the text material and an integration of what the person already knew with the text contents.⁵

The fourth and fifth measures of understanding involved verification of factual data. To determine the fourth measure, we presented individuals with a statement after they had read the text and asked whether the contents of the statement could be correctly inferred from the text contents. There were ten correct and ten incorrect statements. For the principle verification task, the fifth measure of understanding, we asked individuals to compare the historical issue in the present research, the Irish potato famine of the nineteenth century, to four other events that varied in similarity to the potato famine in a surface and/or a deep manner. The stock market crash of 1929 was regarded as not similar in either way, the Black Plague, because of its natural causes and its loss of life, was regarded as having surface similarity, (the loss of life in that event, however, occurred at all levels of society). Recent outbreaks of tuberculosis were seen as having both surface and deep similarity, with disadvantaged people being victimized more than the advantaged. The post-Civil War poll tax was assumed to be similar to the famine at a deep level, since poor populations were the victims in both cases, but dissimilar on the surface, because the tax caused no deaths. We were interested especially in answers to the poll tax item because we assumed that those who observed the similarity between it and the famine would have greater understanding of the events. Individuals were asked to rate each of the four test items on a scale of 1-10 according to how similar its causes were to those of the Irish potato famine.

Rationale and Procedures for the Present Studies

Rationale

A number of efforts have been made to enhance learning and understanding in history. These include improving textbooks,⁶ using multiple documents,⁷ providing quality instruction,⁸ and asking students to generate "their own histories."⁹ Thus, an important question is, What type of manipulations or interventions would be expected to facilitate learning and

understanding? On the basis of the theoretical issues that have been discussed, a goal of an instructional procedure should be to produce a deeper processing of the to-be-learned information, a processing that includes an integration of text contents and the reader's knowledge. The two manipulations of the present research were selected with this goal in mind.

As previously mentioned, one manipulation employed in the present studies was presentation of a textbook chapter as a single text in its standard narrative form and presentation of the same text in a multiple-source format in which the text was divided into eight segments, presented in a random order as separate sources. In both conditions, individuals were asked to read all of the text.

The rationale for this text manipulation was our first hypothesis: that the multiple-segment condition would yield deeper processing than the single-text presentation because the reader in the multiple-segment condition would need to integrate the eight sets of content in order to develop a coherent text structure, while in the standard-text condition the structure was already present in the text.

The other instructional procedure we studied consisted of having different individuals write different types of essays. There were three types in the first study: a narrative, an argumentative essay, and a historical essay, with "history" left undefined. The instructions were identical in all three conditions, except for the insertion of one word, "narrative," "argument," or "history." In some of the other studies, explanatory and summary essays were also requested. The focus of the present discussion is, however, a comparison of the process of writing a narrative versus that of writing an argumentative essay.

A second hypothesis tested by the essay writing manipulation is that writing an argumentative essay would require more and deeper processing of the material than writing a narrative, because the former requires more rearrangement and reorganizing of the text material than the latter. Indeed, extensive processing should be required when the multiple-source text version is read and an argumentative essay is written. The point constituted our third hypothesis: that understanding measures would be of the highest magnitude in the multiple-source, argumentative essay task.

As to the learning measures, which would require only a surface comprehension of the text contents, we hypothesized that learning would not differ among the conditions, since a reasonable surface comprehension was expected in all cases. This constituted our fourth hypothesis.

Materials and Procedures

The historical set of events chosen for the studies was the nineteenth-century Irish potato famine. The text for all of the five studies described the religious, political, sociocultural, and agricultural factors involved in the famine and the resulting population decline in Ireland that occurred in the mid-nineteenth Century.

In all studies, individuals were given two packets. The first contained materials that asked them to read one of the two versions of the text, the single text or the multiple-segment version. The contents of the two versions were identical except for the difference in order and the need to include a few sentences in the multiple-segment condition to provide a sense of flow. These sentences did not contain any causal or explanatory information. The separate "segments" consisted of a map; biographical accounts of King George III and Daniel O'Connell; brief descriptions of the Act of Union, the Act of Emancipation, 1829, and the Great Famine; census data on population size, the death rate, and the rate of emigration between 1800 and 1850; economic statistics on crop prices, rent costs, and distribution of land holdings; and occupational breakdowns for the years between 1800 and 1850. The excerpts were taken from a number of texts and constructed into a textbook-like chapter.¹⁰

When writing the essays, individuals always had access to the text they had read. This procedure was followed so that the essay writing would not be dependent upon memory. In other words, we were interested in how the writer would use the contents of the essay, not in how well the person could remember the text contents in writing the essay.

The second packet contained materials for testing. Depending upon the study, some packets contained a number of completion questions to measure learning and a sentence verification task in which individuals indicated whether each item of twenty appeared in the text contents (ten did and ten did not). The inference and principle verification items, intended to measure understanding, were also included in the packet.

Findings and Their Implications

In general, the most important finding of our studies was that reading from multiple segments combined with writing an argumentative essay yielded deeper understanding of the material than any other condition in

which text format and essay type were manipulated. This result supports Hypothesis 3 and is consistent with the idea that presentation of multiple texts and the writing of an argumentative essay requires more processing than the other conditions. The question of whether the multiple-segment condition produced better understanding than the single-text condition and whether writing an argumentative essay produced better understanding than writing a narrative essay may be answered in the affirmative, although with qualifications.

Regarding the overall organization data, that is, the writing of a listing versus an analytic essay, in general, the multiple-segment condition yielded more analytic essays than the single-text condition. Furthermore, more analytic essays were written in the multiple-segment, argument-essay condition than in any other condition, while more listing essays were written in the single-text, narrative-writing condition. These findings support the general idea that, with respect to understanding, there is a relationship between the way information is presented and the particular task involved in using the information.

With respect to the causal-connective measure, the proportion of causal connectives was greater in the argument condition essay than in the narrative essay condition. While the single-versus-multiple segment variable did not yield a significant difference, the multiple-segment argumentative-essay condition yielded the greatest proportion of causal connectives, up to a third of the connectives used.

These excerpts from one participant's essay are typical of an analytic essay that contained a number of causal connectives.

There are several reasons that led to the sharp decrease of the Irish population between the years 1846–1850. They include the great famine, which had a devastating effect....

Another reason causing the decline of population was the Catholic Emancipation in 1829. George IV granted Catholics the same right to worship as the Protestants, but raised the qualifications for voting. These were based on property value, and this caused the voting population to fall from 100,000 to 16,000. Death rate per 1000 people was a whopping 50, more than double the mark set by the population in the 5 years previous.... Also, the immigration rate in Ireland was off the charts. In 1850 alone, 250,000 people left the Emerald Isle compared to 1838 when only 15,000 left.

The population in Ireland decreased in the years 1846–50 because people were being persecuted for their beliefs or did not have a strong voice in government.

In contrast, the following listing essay lacks causal connectives.

The lives that were being lost somewhat made the significant change in Ireland's population. In 1846 to 1880, it was all about survival for Ireland's population. The overlapping cold and what winters didn't make anything better for Ireland's population. There were people struggling to survive off bad crops. The decrease in average shows that there were less food supplies, and to top it off the crops that were bad were being produced were regularly being taxed. The feeling under the ruling of the United Kingdom may have affected the population of Ireland.

The multiple-segment condition and argument-writing task tended to yield a higher proportion of transformed sentences than the single-text-book condition or the narrative-writing condition. However, the greatest proportion of transformed sentences was obtained when the multiple-segment and argument-writing conditions were combined. The proportion of added sentences varied little according to the essay type or presentation format. However, the proportion of borrowed sentences was greater in the narrative-essay condition than in the argument-essay condition, especially in the single-text condition. This finding supports the idea that the construction of narrative essays involved less processing that was necessary in the argument-essay condition.

In relation to the inference verification task, students who read from multiple segments and wrote argumentative essays were best at recognizing inferences that followed from the text.

The principle verification data, which required the participant to recognize the similarity between the famine and other events, indicated that individuals who wrote argumentative essays from multiple segments were more likely than individuals who wrote narrative essays or who wrote from single texts to rate the poll tax as more similar to the famine than the other events. These results support the idea that understanding is facilitated by writing an argumentative essay from multiple segments.

Turning to the learning measure of recall, which was obtained in one study, the primary finding was that, while participation in the single-text narrative-essay and the multiple-text argumentative-essay conditions did not differ in the number of correct items recalled, they both demonstrated better recall than did participants in the other conditions. This finding suggests that, when reading a narrative text, readers develop a narrative representation of the contents. When they are then asked to write a narrative essay on the subject, a type of resonance occurs, making

the task somewhat easier and reducing the amount of processing required. This congruence also is consistent with the relatively high number of borrowed sentences in this condition.

The sentence verification data, which also are a measure of learning, produced better recognition in the narrative-writing tasks than in the argument-writing task, especially when students wrote from single texts. This suggests that students who wrote narratives from single sources engaged in more extensive processing at the textbase level. In other words, writers of narratives were more concerned with being accurate in relation to the text than in developing their own models of the situation. However, if asked to write an argumentative essay instead of a narrative, individuals need to construct the argument on the basis of a narrative representation, which requires considerable processing of the text content.

In the multiple-segment condition, the reader is able to develop a narrative essay with a relatively small amount of processing. But, if asked to write an argumentative essay, the writer finds the representation narrative of little help; especially in the multiple-segment condition, substantial processing is needed. These notions suggest that individuals, to understand or make sense of events, are predisposed to place them in a chronological, narrative form and that writing an argumentative essay requires deliberation or processing that are not present in the narrative.

We wish to note three other findings. In one study, results supported the account of representation development that we have given. All participants read the multiple-segment text and subsequently rated each sentence of the text for its importance to the overall text contents. Half of the individuals wrote narratives, and half wrote argumentative essays. For each type of essay, half of the participants rated the sentences before writing their essays, and half rated them after writing them. We found that the advantage in understanding in the argument essay condition, compared to the narrative condition, occurred when the ratings in both conditions were performed after the essay writing. However, when the ratings were done before the essay writing, those who wrote argumentative essays demonstrated no advantage with respect to understanding. Specifically, considering the number of causal connectives, the argument condition yielded a mean of 7.8 before the ratings but 5.4 afterward. The narrative essay yielded means of 4.9 and 5.4 for ratings before and after essay writing, respectively. This finding indicates that taking time to rate sentence importance interferes with the construction of an argumentative essay

but not with the construction of a narrative essay. This raises the interesting possibility that argumentative essay writing involves considering the study contents in relation to a standard of importance different from that used in a text-based narrative. In other words, an argument structure may attribute importance based on supporting and opposing reasons to the claim rather than on utility in reconstructing the narrative.

The second and third results were obtained in the fifth study in the series. Two additional procedures were employed in that study that were not part of the other studies. One of these consisted of studying student perception of the narrative- and argumentative-essay tasks. This procedure was employed in order to understand better the differences we had found in essay writing in the earlier studies. Students were asked to state what they understood to be narrative- and argumentative-essay tasks. Their perceptions of the tasks were classified into two categories, using a keyword procedure. One class included essay descriptions using terms such as "story-like," "one-sided," "descriptive," and "writing your opinion." The other class included terms involving presentation of "two sides" or "both sides," "defending," "backing up," "interpreting" facts, or "proving." The former list was regarded as perceiving a more one-sided and descriptive view, a relatively simple view, whereas the second class was regarded as perceiving two sides and the need for evidence, a more complex perception.

The results indicated that individuals who wrote analytic, as opposed to listing, essays were more likely to have complex views of essay writing rather than a simple view. The more complex view was also positively related to having a greater number of transformed sentences in the essays and a greater proportion of correct inference verifications and principle verifications. These results point to the important conclusion that for text understanding to be maximized via writing essays, student knowledge of essay structure and especially of how different viewpoints or perspectives are related and can be incorporated into the essay structure is critical.

The other procedure added to the fifth study involved asking individuals to read two essays about the Irish potato famine. The rationale behind this change is this: The results have indicated that, when a student writes an argument from multiple materials, a causal or complex model that considers multiple sides of an argument is more likely to be developed. From this result it may be inferred that a person who develops such a complex, causal model that incorporates both sides of an issue may also

be better at detecting bias in another account. To test this notion, students were presented with two additional essays after they completed the reading and writing tasks. One essay was an abbreviated version of a lecture given by the Reverend John Hughes at the Broadway Tabernacle on March 20, 1847, and published by the *New York Times* on March 21, 1847. The title of the lecture was "A Lecture on the Irish Question"; in it, Hughes argued that both the English and the Irish were responsible for the effects of the famine. The second document was an editorial by Thomas Campbell Forster, published in the *London Times* on March 15, 1847, which suggested that the Irish had brought on the famine by themselves. The Forster article in particular was quite a biased account of the famine. On a scale of 1–10, students rated their agreement or disagreement with each article.

While we found no significant differences in the level of agreement with the two essays as a function of argumentative- or narrative-essay writing, those students who perceived the writing task as more complex were more likely to disagree with Forster's editorial, suggesting the observation of bias in the essay. We did not find significant differences between the two groups of participants in their ratings of the Hughes paragraph. The bias ratings found in the Forster essay, however, suggest that those individuals who were developing a more complex, causal representation of the famine were more likely to perceive bias in the Forster essay and to disagree with it.

Instructional Considerations

What does this research suggest about instruction in history? A number of points are noted. First, understanding is a function of the nature of the information processing that takes place both during reading and during performance of a task, such as writing. Moreover, the processing that takes place is related to the depth of understanding that is achieved. Learning the text contents *per se* is possible with relatively superficial processing, as well as with deeper processing, but understanding occurs only with deeper processing.

Second, deeper processing is facilitated by the individual's prior knowledge, of the specific topic, related topics, and history in general and by a more advanced level of general information and thinking skills, such

as knowledge of essay structures. Furthermore, such knowledge needs to be utilized in particular contexts in order to facilitate processing.

Third, having individuals use multiple sources—even segments of several secondary accounts—and write argumentative essays constitutes a combination of procedures that helps to maximize processing. When students' perception of the writing task included the need to integrate information from multiple materials into an argument with two sides, the students were more likely to transform and integrate the presented information. Moreover, in doing so, students developed better causal models of the potato famine and a better understanding of its causes, as evidenced by their better performance on the inference verification and principle identification tasks. Furthermore, the development of their own complex causal models allowed students to see the bias in another account.

Other research on learning from multiple materials has also indicated that reading multiple sources can facilitate historical understanding, for example, by encouraging the comparison of contents across texts, making students aware of the importance of source information, and producing the recognition of inconsistencies and biases within texts. Some studies go beyond the present exercise in deliberately involving different perspectives, but the results of using multiple segments of information are clearly important in their own right.¹¹ It is also desirable to have students construct and synthesize their own histories from documents, which can help students develop an understanding that history is more than "someone else's facts."¹² And, especially when they are asked to write problem-based essays from multiple materials as opposed to a historical narrative, students are more likely to compare and cite information from the various sources in their essays.¹³

The present results are also related to the idea of integrative complexity,¹⁴ Measurement of integrative complexity involves a content-free procedure in which a text segment is analyzed in terms of whether it considers two sides of an issue or only one. The measure has been used extensively in political science with documents and speeches, and results generally show that integrative complexity is greater when issues are analyzed more thoroughly, and lower when one-sided and sometimes dogmatic views are stated in the text. This measure fits well with the present findings, since one would expect the perception of the writing task as requiring a two-sided or integrated approach to be related

to greater complexity of material. The measure could conceivably provide a means to evaluate student development in writing argumentative discourse.

NOTES

1. Jennifer Wiley and James F. Voss, "The Effects of Playing Historian upon Learning in History," *Applied Cognitive Psychology*, 10 (1996), pp. 563-572; James F. Voss and Jennifer Wiley, "Developing Understanding While Writing Essays in History," *International Journal of Educational Research*, 27 (1997), pp. 255-265; Jennifer Wiley and James F. Voss, "Constructing Arguments from Multiple Sources," *Journal of Educational Psychology* (1999).
2. Walter Kintsch, *Comprehension* (Cambridge, 1998).
3. Bonnie J. F. Meyer, "Prose Analysis: Purposes, Procedures, and Problems," in Bruce Britton and John Black, eds., *Understanding Expository Text* (Hillsdale, NJ, 1985), pp. 11-64.
4. Stuart Greene, "Students as Authors in the Study of History," in Gaea Leinhardt, Isabel L. Beck, and Catherine Stainton, eds., *Teaching and Learning in History* (Hillsdale, NJ, 1994), pp. 137-170.
5. Carl Bereiter and Marlene Scardamalia, *The Psychology of Written Composition* (Hillsdale, NJ, 1987).
6. Isabel L. Beck and Margaret G. McKeown, "Outcomes of History Instruction: Paste-up Accounts," in Mario Carretero and James F. Voss, eds., *Cognitive and Instructional Processes in History and the Social Sciences* (Hillsdale, NJ, 1994), pp. 237-256.
7. Charles A. Perfetti, M. Anne Britt, and Mara C. Georgi, *Text-Based Learning and Reasoning* (Hillsdale, NJ, 1995).
8. Samuel S. Wineburg and Suzanne M. Wilson, "Subject Matter Knowledge in the Teaching of History," *Advances in Research on Teaching*, 2 (1991), pp. 305-347.
9. Tom Holt, *Thinking Historically* (New York, 1990).
10. J. Mokyr, *Why Ireland Starved: A Quantitative and Analytical History of the Irish Economy, 1800-1850* (London, 1983); J. M. Goldstrom and L. A. Clarkson, eds., *Irish Population, Economy, and Society: Essays in Honour of the Late K. H. Connell* (Oxford, 1981); K. S. Bottingheimer, *Ireland and the Irish* (New York, 1982); R. D. Edwards, and T. D. Williams, *The Great Famine* (New York, 1976); Charles Sidman, "From the Act of Union to the Fall of Parnell," in Harold Orel, ed., *Irish Society and Culture: Aspects of a People's Heritage* (Lawrence, KS, 1976); E. E. R. Green, "The Great Famine," in T. W. Moody and F. X. Martin, eds., *The Course of Irish History* (Cork, 1984).
11. Charles A. Perfetti "Sentences, Individual Differences, and Multiple Texts:

- Three Issues in Text Comprehension," *Discourse Processes*, 23 (1997), pp. 337-355.
12. Holt, *Thinking Historically*.
13. Greene, "Students as Authors."
14. Philip E. Tetlock, "Integrative Complexity of American and Soviet Foreign Policy Rhetoric: A Time Series Analysis," *Journal of Social and Personality Psychology*, 49 (1985), pp. 565-585.