7 Learning from Writing: Study Two

The differences that emerged in the initial case study led us to explore more systematically the effects that various classroom tasks have on learning from text. The second study examined a broader range of tasks and passages and tested the effects over a longer term (one month instead of a few days). We had two primary concerns: (1) to document the longer term effects of writing versus not writing (represented by a read-and-study task), and (2) to explore the effects of writing tasks that require reformulation of new information versus simpler ones that focus on review. For reformulation, we developed tasks requiring analytic writing; for review, we chose two typical approaches, note-taking and answering comprehension questions.

Participants

For this study, we obtained the cooperation of the English department of a local secondary school. A sample of 208 students was drawn from six ninth-grade and six eleventh-grade classes. The students represented the full spectrum of abilities at each grade level, except that classes for English as a second language and classes for the educationally mentally handicapped were excluded from the sample.

Passage Selection

Four passages were selected from high school social studies texts. Two of the passages ("economic expansion" and "the Great Depression") were those used in the exploratory study, and two additional passages were selected for the present study. One of these dealt with political and economic developments in Russia after World War II; the other discussed the influence of science on life in the twentieth century. Though drawn from longer units, all four passages were self-contained and able to stand alone. (See Appendix 2 for synopses of the four passages and their characteristics.)

Study Conditions

Four study tasks were designed for each passage: normal studying, note-taking, comprehension questions, and analytic writing.

Normal studying. Students in the normal studying condition were told simply, "Study the way you normally do to remember the information in the passage." This condition allowed us to examine how students would approach the task when allowed to choose their own methods.

Note-taking. Students in the note-taking condition were told, "Take notes to help you learn the information in the passage." This is a review activity that allows the students to concentrate on the material they consider most relevant.

Comprehension questions. For the comprehension-question condition, we designed a series of short-answer questions similar to those that students encounter in workbook study guides and teacher-made dittos. Review activities of this sort focus the students' attention on specific aspects of the passages. For each of the four passages, twenty questions were devised and divided equally among textually explicit and textually implicit questions. Sample items about "economic expansion" follow:

> Please answer the following questions as you would answer questions for a homework assignment.
> *Economic Expansion:*
> What were the major manufacturing industries in the United States at the turn of the century?
> What did profits on goods, bank loans, and foreign investments have in common?

Analytic writing. In the analytic-writing assignments, the students were asked to reformulate and extend the material from the passages as they developed evidence to support a particular interpretation or point of view. For "economic expansion," the students were asked to respond to the following question:

> Given what you learned from the passage, what do you feel were the two or three most important reasons for industrial growth in the late nineteenth and early twentieth centuries? Explain the reasons for your choices.

Measures

Three instruments were designed to examine what students had learned in the process of reading and studying the passages. These measures are described below.

Topic Knowledge

Langer's (1980, 1981, 1982, 1984b, 1984c) measure of passage-specific knowledge was used to measure students' knowledge of the topic and how it changed as a result of particular study activities. Students were asked to provide written free-association responses to five key concepts drawn from the top half of the content hierarchy in each passage (see Meyer, 1975, 1981). An unrelated concept (dog) was used as a practice item before the five words were presented. Practice exercises were given orally, and students were paced through the free-association task one concept at a time. Sufficient space was left between concepts so that the students could provide as many associations as possible.

The measure was scored to reflect a combination of the amount (breadth) and organization (depth) of passage-relevant information reflected in the free associations, using procedures developed by Langer (1980, 1984b, 1984c; Langer and Nicholich, 1981). For each concept word in the knowledge measure, each free association was scored as indicating (1) peripheral knowledge of the concept, (2) concrete understanding (such as examples, attributes, defining characteristics), or (3) abstract understanding (such as superordinate concepts, definitions). Ratings reflecting levels 2 and 3 were then summed across concepts and raters to derive a total score for each passage. Interrater reliability for the total score (estimated using the Spearman-Brown formula) was .875. The test-retest correlation was .712 after four weeks and an intervening treatment period.

The measure was administered three times: before the students had read the passage, immediately after reading it, and four weeks later. As a pretest, this measure reflects students' prior knowledge of the topics they read about; changes between the pretest and the post-test provide a measure of what the students learned as a result of the reading and study activities.

Passage Comprehension

A twenty-item multiple-choice test was constructed for each passage to measure overall comprehension. Eight items required a simple report of information from the passage, eight required the student to construct relationships among items of information in the passage, and four required drawing generalizations that extended beyond the passage. To ensure that items and distractors were functioning as intended, the items for each test were developed through a cycle of pilot testing that included interviews exploring the participants' reasons for their answers. The twenty items for each passage were randomly ordered

and administered once, four weeks after the initial read-and-study tasks. The multiple-choice items were scored right or wrong and summed to give the total number correct (out of twenty) for each passage.

Application of New Information

The final measure was an extended essay that required students to orchestrate what they had learned in a coherent argument based on information from the original reading. Though requesting the same type of writing as the analytic-writing study condition, the format of the prompt and specific topic differed in each case. The essay was administered at the four-week post-test. For example, the instructions for "economic expansion" read:

> Write an essay based on what you learned from the reading on economic expansion. Use the title, "Causes and Effects of Industrial Growth at the Turn of the Century." Be certain to support the points you make.

The essays were ranked by two independent raters on the basis of overall coherence and the structure of the argument developed, rather than on the conventions of standard written English. The essays from the four passages were scored on a single scale from best to worst. Tables of the normal distribution were used to convert each rater's scores to a normally distributed scale ranging from 22 (best) to 1 (worst). Scores for the two raters were then summed to yield an essay-quality score with a sample mean of 23.2 and standard deviation of 7.5. Interrater reliability for the total score (estimated using the Spearman-Brown formula) was .94. This procedure, though obviously not feasible in larger scale assessments, provided much better discrimination among essays than would have been gained from more common 4-point or 6-point holistic or general impression rating scales.

Procedures

Separate but overlapping sets of three passages were used at each grade level (ninth and eleventh). The "Great Depression" was used only with eleventh graders; "twentieth-century science" only at grade nine. At each grade level, two classes were assigned at random to each passage. (Passages were assigned by class to simplify administration of passage-specific measures.) Study packets were assembled so that students within classes were randomly assigned to one of the four study conditions.

During the first day of the study, the students completed the passage-specific knowledge measure, followed by a packet containing (1) general directions to read the passage and then to complete the task that followed, (2) the reading passage, and (3) directions for the study task. Ten minutes before the end of class, the passages and study packets were collected and the passage-specific knowledge measure was re-administered. Students had seven minutes for each administration of the knowledge measure, and thirty-five minutes to read the passage and complete the study task. The study tasks thus functioned as post-reading activities, with the reading passage available while the study tasks were completed.

Exactly four weeks later, all classes completed the three measures of learning. The passage-specific knowledge measure was given first, followed by the essay test focusing on comprehension of relationships within the original passage. The multiple-choice comprehension tests were administered last so that the questions and answers would not provide students with additional information to draw upon in completing the other measures. Again, all measures were completed within a single class period, with seven minutes for the passage-specific knowledge measure, twenty minutes for the essay, and twenty minutes for the comprehension test. (At this session, students did not have any of the materials from the original study session available to them.)

Responses to the Study Tasks

The pretest measure of passage-specific knowledge provides a test of the initial comparability of the four groups. The relevant results, summarized in table 10, indicate that students in the four study conditions did differ somewhat in the extent of their initial passage-specific knowledge. The normal studying and the note-taking groups had somewhat higher initial knowledge of the topics discussed in the passages they read. Grade level differences were not significant because of the use of an additional, easier passage with the ninth graders and an additional, harder passage with the eleventh graders. (The eleventh graders had significantly greater passage-specific knowledge than did the ninth graders for the two passages given to both grade levels, $F(1;202) = 14.56$, $p < .001$.) Because of the initial differences in passage-specific knowledge, the analyses that follow use initial passage-specific knowledge as a covariate in order to provide a statistical adjustment for the initial group differences.

The amount that students wrote in response to each study task provides one indication of the amount of cognitive effort that they

Table 10

Characteristics of Student Performance on Selected Study Tasks

	Adjusted Means			
	Essay (n = 53)	Comprehension Questions (n = 47)	Note-taking (n = 54)	Normal Studying (n = 54)
Pretest passage knowledge				
Grade 9	11.4	10.7	11.9	13.2
Grade 11	12.8	10.2	16.3	13.1
	(Pooled within-cell SD = 8.15)			
Words written during task				
Grade 9	99.4	114.2	101.1	21.7
Grade 11	123.5	94.5	155.1	54.1
	(Pooled within-cell SD = 52.99)			

Analysis of Variance

	Initial Passage Knowledge			Task Words		
Effects	df	F	p	df	F	p
Task	3	2.62	.052	3	21.52	.001
Grade	1	2.41	.122	1	0.38	n.s.
Passage	3	31.15	.001	3	6.02	.001
Task × passage	9	1.47	.160	9	0.83	n.s.
Task × grade	3	0.30	n.s.	3	0.28	n.s.
Passage × grade	1	1.84	.176	1	3.94	.049
Task × passage × grade	3	0.71	n.s.	3	0.14	n.s.
Error	184			184		

put into each task. In turn, we would expect the amount of effort to be related to the amount of learning that resulted. These data are also summarized in table 10. As would be expected, the eleventh graders wrote more than the ninth graders, and the normal studying group wrote on average less than the groups that were specifically asked to write. At grade nine, the three types of writing tasks produced relatively similar amounts of writing, but at grade eleven they diverged somewhat, with comprehension questions producing the least writing and note-taking the most.

Table 11

Multiple-Choice Comprehension at Four Weeks

	Adjusted Means			
	Essay (n = 53)	Comprehension Questions (n = 47)	Note-taking (n = 54)	Normal Studying (n = 54)
Grade 9	9.0	9.0	9.2	8.7
Grade 11	9.6	10.8	10.4	10.6
	(Pooled within-cell SD = 2.63)			

Analysis of Variance[a]

Effects	df	F	p
Task			
Linear	1	0.41	n.s.
Deviations from linear	2	0.05	n.s.
Grade	1	8.78	.003
Passage	3	29.92	.001
Task × passage	9	1.19	n.s.
Task × grade	3	1.64	.181
Passage × grade	1	0.01	n.s.
Task × passage × grade	3	0.88	n.s.
Covariate	1	19.37	.001
Error	183		

[a] Task × passage × grade, covaried on pretest passage knowledge.

Effects of Study Tasks on Learning

To what extent did the different study conditions lead to different effects on learning? Results for the multiple-choice comprehension test are summarized in table 11. The effects of most interest, those involving tasks, reflect differences among the four study conditions. These effects are partitioned into linear and deviations from linear effects in order to reflect the ordering of the four tasks from the most focused (essay) to the least focused (normal studying). The results indicate that there were no differences among the four study tasks in their effects on the multiple-choice comprehension task given at the four-week post-test, though passage and grade level both had a significant effect on post-test performance.

Results for the passage-specific knowledge measure are summarized in table 12. They indicate that the essay-writing group scored consistently lower than groups in the other conditions, and the normal studying group did consistently better at both the immediate and four-week post-test. The magnitude of this difference was greatly reduced at four weeks, a trend that is reflected in the task (linear) × time interaction ($p < .113$). The scores at the immediate post-test suggest that the essay task focused students' attention on a narrower range of information in the passage, thus providing them with fewer specific associations for the passage knowledge measure and leading to lower passage-specific knowledge scores. In contrast, the normal studying and note-taking conditions may have led students to distribute their attention more evenly over information in the passage as a whole, providing a broader base of associations on which they could draw and thus higher passage-specific knowledge scores. On the other hand, groups that showed the greatest immediate gains also showed the greatest falling off between the immediate and four-week post-tests: the decline from immediate to four-week post-test averaged 3 percent for the essay-writing group, 6.8 percent for the comprehension-question group, 10.5 percent for note-taking, and 11.5 percent for normal studying. Thus the normal studying and note-taking conditions seem to have led to an initial greater breadth of knowledge, but this knowledge was not retained as well.

The third measure from the four-week post-test was the quality of the essay that required students to apply what they remembered from the passage in support of an argument or interpretation. For this measure, consistent task differences again appeared, but in the opposite direction from those that occurred for the passage knowledge measure: the essay and comprehension-question conditions were consistently superior to the normal studying and assigned note-taking groups (table 13). At grade nine, the essay-writing group performed better than the comprehension-question group, but at grade eleven the performance of the two groups was indistinguishable.

It is interesting that the essay scores showed task differences favoring the more focused writing conditions at four weeks even though the other measures did not. The essay task differed in three important ways from the other two outcome measures: it provided fewer cues to recall, required orchestration of relationships among the information that was remembered, and could be completed successfully using a narrower selection of information from the original passage. This finding suggests that the two study conditions requiring the most focused writing — the essay and comprehension-question study tasks —

Table 12

Passage-Specific Knowledge: Immediate and at Four Weeks

	Adjusted Means			
	Essay (n = 51)	Comprehension Questions (n = 45)	Assigned Notes (n = 53)	Normal Studying (n = 53)
Immediate post-test				
Grade 9	17.8	19.2	22.0	22.8
Grade 11	20.9	24.8	22.6	27.8
		(Pooled within-cell SD = 7.74)		
Four-week post-test				
Grade 9	18.3	18.1	20.1	20.9
Grade 11	19.2	22.9	19.8	23.9
		(Pooled within-cell SD = 8.09)		

Repeated Measures Analysis of Variance[a]

Effects	df	F	p
Between subjects			
Task	2	2.40	.070
Linear	1	6.47	.012
Deviations from linear	2	0.35	n.s.
Passage	3	5.24	.002
Grade	1	0.82	n.s.
Task × passage	9	0.66	n.s.
Task × grade	3	1.02	n.s.
Passage × grade	1	16.41	.001
Task × passage × grade	3	1.40	n.s.
Covariate	1	100.78	.001
Error	177		
Within subjects			
Time	1	10.92	.001
Task × time	3	0.99	n.s.
Linear	1	2.54	.113
Deviations from linear	2	0.21	n.s.
Passage × time	3	9.81	.001
Grade × time	1	2.88	.091
Task × passage × time	9	0.18	n.s.
Task × grade × time	3	0.31	n.s.
Passage × grade × time	1	2.47	.118
Task × passage × grade × time	3	0.45	n.s.
Error	178		

[a] Task × passage × grade × time, with pretest passage knowledge as a covariate.

Table 13

Essay Quality at Four Weeks

	Adjusted Means			
	Essay (n = 46)	Comprehension Questions (n = 42)	Note-taking (n = 49)	Normal Studying (n = 48)
Grade 9	23.5	21.6	20.6	20.8
Grade 11	27.2	27.6	23.3	24.0
	(Pooled within-cell SD = 6.87)			

Analysis of Variance[a]

Effects	df	F	p
Task			
Linear	1	4.26	.041
Deviations from linear	2	0.95	n.s.
Grade	1	1.33	n.s.
Passage	3	3.67	.014
Task × passage	9	1.29	n.s.
Task × grade	3	0.48	n.s.
Passage × grade	1	0.55	n.s.
Task × passage × grade	3	0.23	n.s.
Covariate	1	8.36	.004
Error	160		

[a] Task × passage × grade, covaried on pretest passage knowledge.

may have led to a deeper understanding of a narrower body of information than did the note-taking and normal studying tasks.

Effect of Amount Written on Post-test Performance

When we examined the number of words written during the study task, we hypothesized that the number would reflect the cognitive effort students put into their writing. That is, we would also expect that writing *more* would be related to better post-test performance, whatever particular writing task a student may have been assigned. To examine this hypothesis, we can look at the relationship between the amount written during the study task and post-test performance after accounting for all of the other factors and covariates in the model (passage, task, grade level, and pretest passage-specific knowledge).

Table 14 presents the relevant pooled within-cell correlations for each study condition separately and for all four groups pooled. In

Table 14

Relationships between Words Written during Study Task and Post-test Performance, Adjusted for Task, Grade, Passage, and Pretest Passage Knowledge

	Adjusted Within-Cell Correlation with Words Written during Study (*df*)				
Measures	Essay	Comprehension Questions	Note-taking	Normal Studying	All
Essay quality	.451** (39)	.290* (35)	.014 (42)	.273* (41)	.240*** (160)
Passage knowledge					
Immediate	.301* (44)	.259* (40)	.158 (47)	.283* (46)	.234*** (180)
Four weeks	.415** (46)	.525*** (38)	.082 (46)	.190 (47)	.292*** (180)
Multiple-choice Comprehension	.023 (46)	.314* (40)	−.038 (47)	−.104 (47)	.026 (183)

* $p < .05$
** $p < .01$
*** $p < .001$

general, performance on the multiple-choice comprehension test showed little relationship to the amount written for any of the groups except for the one that had answered similar questions during the study period. For the other measures, the amount written in response to either the essay task or the comprehension questions was positively and significantly related to post-test performance. Interestingly, for the passage knowledge measure these effects are stronger at four weeks than at the immediate post-test — a reflection perhaps of recency effects in initial responses to that measure. The consistent positive relationships for students in the essay-writing and comprehension-question study tasks indicate that, for these types of writing at least, the writing process itself may be directly related to the learning that results. For the other two conditions, note-taking and normal studying, the effects on learning may be associated with spending time with the material, whether or not much writing is involved. It is important to remember that the correlations have been corrected for pretest performance; they are not simply the result of good students doing better in everything including writing more.

Discussion

Results from this study are interesting but complicated. Rather than showing general effects, the results show that task differences favoring writing emerge only on the more complex and time consuming of the outcome measures, the essay requiring students to use what they had learned in order to mount an argument of their own. The other measures, which may have tapped a broader spectrum of remembered information, either show no differences or yield results favoring the normal studying and note-taking conditions.

The superior performance of the two focused writing groups on the four-week essay is encouraging, given our general hypotheses about the relationships between writing and reasoning. On the other hand, the effects are relatively small, and the differences among the various conditions are difficult to untangle. Results from the immediate post-test using the topic-knowledge measure suggest that the essay task may have focused students' attention on a narrower band of information, though by four weeks the advantage to the other conditions had been considerably reduced. The evidence from the within-cell correlation measures also suggests that there may be a relationship between what was written about and what was remembered, at least when students were completing focused writing tasks. At the least, writing more seemed to be related to how much was remembered later.

A third study, presented in chapter 8, was designed to pursue some of the questions raised by the one presented here. With more focused measures of outcomes, would clearer differences be discernable among various types of writing tasks? Could behavior during the study task, reflected here only in the number of words written during the treatment, be more directly traceable to post-test performance? If other types of focused writing were required, would they yield outcomes comparable to those for the essay-writing and comprehension-question conditions in the second study?