Facilitating Text Learning

Andrew D. Katayama
West Virginia University

Daniel H. Robinson
University of Texas at Austin

Kenneth A. Kiewra
University of Nebraska-Lincoln

Nelson Dubois
SUNY-Oneonta

David Jonassen
University of Missouri-Columbia

Three experiments were conducted to investigate whether graphic organizers (GOs) were more effective than concept maps (CMs) in facilitating text learning. Experiment 1 was a replication of Robinson and Kiewra's (1995) second experiment except that only GO and CM conditions were included. Experiment 2 used a different text, and separate GO, CM, and tests constructors to reduce experimenter bias. Experiment 3 replicated Experiment 2 except that the GO constructor also constructed the tests. Results showed that GOs only "facilitate" text learning when they are constructed by someone who also constructs the tests. The implications of this study are that adjunct displays will probably not improve classroom test performance unless the test measures the type of learning the displays are intended to facilitate.

Biases with Adjunct Display Construction and Testing

The authors of this paper share a common interest in how spatial "adjunct" displays can be used to facilitate classroom learning. Each has conducted research that has shown spatial displays (e.g., Jonassen, 1994; Katayama & Robinson, 2000; Kiewra & Dubois, 1998) to be more effective than expository text in helping students learn. However, there has been little research examining what types of spatial displays are most effective. This issue provided the motivation for the current set of studies.

Two spatial displays that have received considerable attention over the past twenty years are graphic organizers (GOs) and concept maps (CMs). Both types of displays were developed to communicate concept relations. The key difference between the two is how they communicate the relations. CMs use a node-link format where names of concepts are enclosed within borders (e.g., ovals, squares) and are connected to each other with lines called links. These links are labeled with a word or words explicitly describing the relation. GOs, on the other hand, simply use spatial arrangement of concepts to communicate their relations, without "spelling out" the relations for the student. For example, by placing the word "whales" above the words "right whale" and "baleen whale," a hierarchical relation is communicated.

CMs had their origin in research conducted by Novak (Novak, 1990). In his earliest work, Novak did not label the links "but it soon became clear that this was essential to represent concept/propositional meanings in an explicit hierarchical framework" (Novak, 1990, p. 938). CMs were initially intended as a learning activity where students themselves construct the maps. This activity is called "concept mapping." Concept mapping is also known as "knowledge mapping" due to the work of Dansereau and his colleagues. Some recent textbooks (e.g., Dembo, 1994;

Kiewra & DuBois, 1998; Seifert, 1999; Woolfolk, 2000), however, have study guides that include CMs as adjunct displays to be studied along with the text.

GOs were first developed by Barron (1969) and by Earle (1969) to address problems associated with Ausubel's (1968) advance organizer. They proposed that a graphic display of words showing a hierarchical organization of important concepts would serve students better than would a written paragraph providing an overview of what is to come. This "structured overview" differed from an advance organizer in its ability to illustrate relations among key concepts found in text due to its spatial format. The term "structured overview" was eventually changed to "graphic organizer" as the instructional position was switched from prereading to postreading (an overview is usually presented before the content).

Robinson and Kiewra (1995) conducted a study where they compared the relative effectiveness of GOs and outlines when studied as a set of referenced figures with a chapter-length text. They found that a set of GOs was more effective than outlines for learning concept relations and applying text knowledge. After they had completed the study, the authors asked a third person to construct some concept maps from the same text. In an attempt to see if the CMs would be as effective as the GOs, the authors of the present study replicated the second experiment except that only GO and CM conditions were used.

More recently, Katayama and Robinson (2000) had found similar results in that GOs were more effective as study notes than outlines (OLs) on transfer (application) tests. They also found that study notes in a partially completed format (half-empty) were superior to complete notes (e.g., ones that instructors pass out in class) on the transfer tests. It was concluded that GOs are effective when students construct them from a partial format.

Experiment 1

Method

Subjects and Design

Twenty-eight students enrolled in an undergraduate educational psychology course were assigned randomly to one of two study materials conditions (text-plus knowledge maps or text-plus GOs). Testing took place in a typical university classroom.

Materials

Materials included the text, the adjunct displays (CMs and GOs), tests, and a questionnaire. A text of about 6,500 words (12 pages, single-spaced) was used. Most of the text was taken from a chapter on abnormal behavior in an undergraduate introductory psychology textbook (Davidoff, 1976). The section on psychopathic behavior was taken from a similar textbook (Rubin & McNeil, 1981). The seven GOs (six matrices and one tree diagram) were constructed by one of the authors, whereas the six CMs were constructed by an outside consultant. The GO and CM displays were not informationally equivalent to each other due to differences in what each constructor viewed as being important.

Tests measured facts, relations, and application. The fact test contained 30 multiple-choice items consisting of 15 non-represented fact items (odd numbers) and 15 represented fact items (even numbers). Non-represented fact items required knowledge of facts presented in the text that were not presented in any of the adjunct displays. Represented fact items required knowledge of facts that were presented in both the text and the adjunct displays. The key difference between the non-represented and represented fact items was that represented fact items required knowledge of a defining characteristic of a main concept (appeared as a heading in the text), whereas non-represented fact items required knowledge of either a nondefining characteristic of a main concept or information unrelated to a main concept. This distinction is similar to ones made by Levin, Bender, and Pressley (1979) regarding central versus peripheral text information, and by Levin and Berry (1980) regarding more versus less important text information. The hierarchical relations test and essay test measured relational learning. The hierarchical relations test contained 21 cued-recall items requiring knowledge of text structure. The test was printed over four pages with items arranged so they would not provide clues to other items on the same page or succeeding pages. The essay test contained two items that required knowledge of coordinate relations. Students were asked to compare and contrast coordinate concepts. The application test contained 15 matching items that required identification of disorders, given novel examples of symptoms. Names of all 16 disorders mentioned in the text appeared in an alphabetical list and students were instructed that some names may have been used more than once or not at all. Examples of each test item appear in Figure 1.

A six-item questionnaire was constructed to examine how adjunct displays differentially affect study behaviors. It contained multiple-choice questions that asked students: (a) how "reader friendly" the text was; (b) how "reader friendly" the figures were (students who studied only text were instructed to ignore this question); (c) whether 60 minutes was enough time to read the text; (d) whether 15 minutes was enough time to study the figures (also ignored by text only group); (e) how much effort students put into learning the information; and (f) how interesting the information was. Table 1presents the frequency of responses for the "reader friendliness" items on the questionnaire.

Procedure

Sessions were conducted in two, one-hour periods on separate days. On the first day, students had one hour to read and study their materials. One day later, students studied their materials for an additional 15 minutes. After that, the

Non-represented Facts

Hebephrenic schizophrenics often talk in incoherent sentences referred to as

a. waxy flexibility.
c. delusions of grandeur.
b. word salads.
d. free floating.

Represented Facts

Feelings of hopelessness, guilt, fatigue, and loss of interest in favorite activities are characteristic of

a. simple schizophrenia
c. depressive neurosis
b. functional psychoses
d. anxiety neurosis

Hierarchical Relations

List the four types of neuroses described in the text.

Essay

Discuss as many differences and similarities as you can among neuroses, psychoses (e.g., schizophrenia), and personality disorders (e.g., psychopathic behavior).

Application

Robin, an adult woman, lived at home under the care of her parents. Sometimes she acted normal, but most of the time she was either sitting in the corner of her bedroom, staring blankly at the wall, or else she was scurrying around the house chattering about various topics.

Figure 1. Examples of test items

materials were collected and the questionnaire wasdistributed, which students completed in a few minutes. Then students completed a series of short tests, one at a time, so that they could not return to complete an earlier test based upon knowledge gained from subsequent tests.

Scoring

All tests were scored without knowledge of group affiliation (tests were previously coded). The three objective tests (fact, structure, and application) were scored in accord with predetermined keys with a maximum score equal to the number of items. A liberal scoring criterion was used for the structure test in that students received credit even if they misspelled or provided a fragment of the name of a disorder. The fact test yielded two scores: a non-represented fact score, and a represented fact score.

The essay test yielded two scores: a relation's score and a contrasting premises score. Relation's scores were determined by the number of between-concept relations stated, whereas contrasting premises scores were determined by the sequence in which the relations were stated. A relation was defined as correctly describing two concepts' values along the same attribute. A contrasting premise was defined as a relation in which the two descriptions were not separated by other attribute descriptions. Thus, students' contrasting premises scores could not be higher than their relation's scores. For example, if a student wrote, "process schizophrenia develops gradually over many years, is quite severe, and has a low recovery rate" and later reported that "reactive schizophrenia is triggered suddenly by stress, is not severe, and has a high recovery rate," he/she would receive a score of six relations and zero contrasting premises because each relation was separated by other attribute descriptions. On the other hand, the statement, "process schizophrenia develops over many years whereas reactive schizophrenia develops suddenly, process schizophrenia is quite severe whereas reactive schizophrenia is less severe, and process schizophrenia has a low recovery rate whereas reactive schizophrenia has a higher recovery rate," would be scored as six relations and three contrasting premises. Responses in the form of a GO or CM were not awarded

points for contrasting premises.

Results and Discussion

Table 2 presents means and standard deviations for the six dependent measures. All statistical tests were conducted at the $p < .05$ level of significance, unless otherwise noted. A 2 by 2 mixed-model analysis of variance (ANOVA) was conducted on the 30-item fact test with fact type (represented or non-represented) treated as a within-subjects factor. Although neither of the main effects for study materials or fact type were significant, $F(1, 26) = .01$, $MSE = 8.81$, and $F(1, 26) = 1.39$, $MSE = 2.52$, respectively; there was a significant study materials by fact type interaction effect, $F(1, 26) = 4.80$. Tests of simple effects were used to follow up this interaction. The CM group scored higher on the non-represented facts measure than they did on the represented facts measure, whereas the GO group performed the same on both measures.

Independent *t*-tests were conducted on the 21-item hierarchical relations test and the 15-item application test. The GO group scored higher than the CM group on the hierarchical relations test, $t(26) = 2.68$, $SE = 1.57$; and both groups scored the same on the application test, $t(26) = -.09$, $SE = 1.59$. Because the two essay scores (relations and contrasting premises) were not independent (a contrasting premise required a correctly stated relation), a multivariate analysis of variance (MANOVA) was conducted. The MANOVA was not significant, Wilks' lambda = .89, $F(2, 25) = 1.55$.

The ordinal variables, responses to the six multiple-choice items, were analyzed using separate Mann-Whitney *U* tests. The only significant difference was between the GO and CM groups in how they perceived the displays' "reader friendliness," $U = 19.5$. GOs (mean rank = 8.89) were rated more reader friendly than CMs (mean rank = 19.50).

The GO and CM conditions differed on three of the measures. Perhaps most interesting of these was the students' perception that GOs were more reader friendly than CMs. These students had not previously been exposed to either type of display so this difference in perceptions is probably reliable. However, the study materials by fact type interaction was probably due to the fact that all of the represented information was referenced in the GOs but not in the CMs. Likewise, the GO group's superiority on the hierarchical relations test may also have simply been due to the fact that all of this information was referenced in the GOs and not in the CMs. In other words, the two observed test advantages for the GO group could simply have been a function of the tests being based on information presented in the GOs.

Most spatial display studies have involved providing students with researcher-constructed displays and then having them take researcher-constructed tests. Experiment 1 was certainly not a "fair" test of the effectiveness of concept maps because the same person who had constructed the GOs had constructed the tests. In other words, there may have been some experimental bias favoring the GOs. One way to eliminate this bias would be to have a different person construct the tests. This would also better simulate a classroom situation in which the teacher, not the adjunct displays author, would construct the tests. Finally, to further control for experimenter bias, different persons should construct the adjunct displays. These issues were pursued in Experiment 2.

Experiment 2

Method

Subjects and Design

Fifty-two different students enrolled in an undergraduate

Table 1
Frequencies for Responses to Study Materials Item on the Questionnaire

	How "reader friendly" were the study materials?	
	Knowledge Maps	Graphic Organizers
very easy to understand	1	9
fairly easy to understand	4	5
somewhat difficult to understand	2	0
very difficult to understand	6	0
Total	13	14

Table 2
Means and standard deviations for the three groups on the teacher made tests in Experiment 1.

		Study Materials	
Test		Knowledge Maps	GOs
Non-represented Facts	*M*	8.50	2.88
	SD	2.07	1.95
Represented Facts	*M*	7.07	8.07
	SD	3.00	2.37
Hierarchical Relations	*M*	8.50	12.71
	SD	3.06	5.03
Coordinate Relations	*M*	1.79	3.57
	SD	1.53	3.96
Contrasting Premises	*M*	1.36	3.29
	SD	1.45	3.87
Application	M	7.43	7.29
	SD	4.40	3.58
Total	*M*	34.64	42.57
	SD	10.80	16.75
	n	14	14

educational psychology course were randomly assigned to one of three study materials conditions (text-plus-concept map, text-plus-GOs, or text-only). Testing took place in a typical university classroom.

Materials

Materials included the text, the adjunct displays (concept map and GOs), and the tests. A text of about 5,000 words (14 pages, single-spaced) was used that was taken from a chapter on classroom measurement in an undergraduate introductory educational psychology textbook (Dembo, 1994). The adjunct displays were constructed by three noted experts in adjunct displays: Dave Jonassen at the formerly of Pennsylvania State University, Ken Kiewra at the University of Nebraska-Lincoln, and Nelson Dubois formerly of SUNY-Oneonta. All three experts received a copy of the text and the following instructions: "You should create some displays that you feel will help students best learn the text. Remember that students will not be forced to study the displays; they will simply be told that a set of study materials is available for their use. You may only use information that appears in the text to construct your displays from (this is to ensure a somewhat fair playing ground)." One expert (in concept mapping) constructed a single concept map, whereas two other experts (in graphic organizers) constructed a set of eight GOs collaboratively.

A fourth expert (in tests and measurement) constructed the test. This expert also received a copy of the text and the following instructions: "You should create a test or tests that you feel will adequately measure whether students learned the information presented in the text. Perhaps it would be best to design a test or tests that measures a number of different types of information contained in the text." The test consisted of short answer, multiple choice, table, essay, and tree items.

Because this was designed to be more of an objective study that was not confounded by experimenter biases, the participants were asked to not discuss their tasks with the other participants. All materials were sent to an independent evaluator at an entirely different institution, who served as the experimenter.

Procedure

Sessions were conducted in a 45-minute period combined with a 35-minute period one day later. First, students had 45 minutes to read and study their materials. Two days later, students studied their materials for an additional 10 minutes. After that, the materials were collected and then students completed the tests in approximately 25 minutes.

Scoring

All tests were scored by the researchers without knowledge of group affiliation (tests were previously coded). All tests were scored in accordance to a predetermined key supplied by Armbruster.

Results and Discussion

Table 3 presents means and standard deviations for the six dependent measures. A multivariate ANOVA was conducted on the entire test scores first, indicating no significant difference among the three groups, $F(10, 90) = 1.02$, Wilk's lambda = .81, $p = .44$. Oneway ANOVAs were also conducted on each of the five types of items to see if the groups may have differed in their performance on a particular measure. Again, there were no significant differences among the three treatment groups on any of these dependent measures. In fact, the only difference that was marginally significant was on the short answer test, $F(2, 49) = 2.81$, $MSE = 6.97$, $p = .07$) favored the text-only group.

The null results of Experiment 2 are not surprising when one considers the "blindness" between the spatial display constructors and the test constructor. Unfortunately, this experiment probably simulates what actually happens in real classrooms when instructional designers and test constructors (i.e., teachers), armed with good intentions, fail to agree on what the student should be learning. Perhaps more disturbing are the limitations, particularly concerning external validity, of previous research on spatial displays. It is possible that spatial displays are only effective when students take researcher-constructed tests. Experiment 3 was designed to investigate this issue.

Experiment 3

The three adjunct displays experts were all informed of the null results of Experiment 2 and were asked to design a short test that they believed would more accurately assess the type of text learning their adjunct displays were intended to facilitate. Unfortunately, only two of the three were able to construct a test for the third experiment. We decided to replicate Experiment 2 using their test.

Method

Subjects and Design

Forty-six different students enrolled in an undergraduate educational psychology course were randomly assigned to one of three study materials conditions (text-plus-concept map, text-plus-GOs, or text-only). Testing took place in a typical university classroom.

Materials

The text and adjunct displays were the same as those used in Experiment 2. The test consisted of 33 fact items and five concept items. All materials were again sent to the experimenter who administered and proctored the tests.

Table 3
Means and standard deviations for the three groups on the third party tests in Experiment 2.

		Study Materials Groups		
Test		Text-only	GOs	Concept Map
Short Answer	*M*	5.00	3.72	2.88
	SD	3.27	2.40	2.03
Multiple-Choice	*M*	1.50	1.50	1.75
	SD	1.10	1.10	0.93
Tree	*M*	3.11	3.39	2.94
	SD	1.57	2.03	2.27
Table	*M*	3.56	3.61	3.50
	SD	1.92	1.72	1.71
Essay	*M*	2.44	2.83	2.69
	SD	2.64	2.28	2.12
Total	*M*	15.50	15.06	13.75
	SD	7.17	6.87	5.88
	n	18	18	16

Procedure

The procedure was identical to that used in Experiment 2, except that students completed the tests in about 20 minutes.

Scoring

All tests were scored without knowledge of group affiliation (tests were previously coded). All tests were scored in accord with a predetermined key supplied by two of the experts.

Results and Discussion

Table 4 presents means and standard deviations for the two dependent measures. Separate univariate analyses of variance were conducted on the outcome measures. There was a significant difference among the groups on the fact test, $F(2,43) = 4.68$, $MSE = 19.40$, $p < .05$. A Fisher LSD test indicated that students who studied GOs scored higher than those who studied only text. There was also a significant effect on the concept test, $F(2,43) = 3.93$, $MSE = 3.71$, $p < .05$. Once again, the GO group outperformed the text-only group.

Summary

In all three experiments, the results indicate the relationship between the person who constructs the adjunct display(s) and the test constructor has an influence over the performance on the tests. This is to suggest that when the test constructor and the adjunct displays constructor (regardless if they are GOs or CMs) are the same, the students have an advantage when using that particular display. This was apparent in all three experiments. In experiment 1, the students using GOs outperformed those with CMs. The same expert who constructed the tests constructed the GOs. In experiment 2 we attempted to control for any possible researcher constructed test biases by having different people construct the adjunct displays (GOs by Kiewra and Dubois, CMs by Jonassen) and had a yet another person construct a test (by Armbruster) based on the text. The results from this experiment yielded no differences among the adjunct displays, but it did indicate a marginal difference favoring the text. As we observed this relationship between researcher constructed displays and tests, we wanted to further investigate this pattern of results in a third experiment. This experiment used tests constructed by two experts in graphic organizers. As we predicted, the students in the GO group outperformed the text-only group.

As a classroom precaution, depending on who is constructing the tests, students who study either adjunct display may not perform any better than with text alone. That is to say if teachers use pre-formatted tests (e.g., those from test banks), using a particular adjunct display may not enhance their students' performance. On the other hand, if teachers construct a particular adjunct display for their students to study from, and also construct the tests, then there is a good probability that those students will demonstrate improved test performance than just using the text alone to study.

Conclusions and Recommendations

The authors of the present studies would like to

Table 4
Means and standard deviations for the three groups on the Kiewra/Dubois tests in Experiment 3

		Study Materials Groups		
Test		Text-only	GOs	Concept Map
Fact	*M*	14.75	19.53	16.40
	SD	3.15	6.51	2.56
Concept	*M*	4.00	5.93	4.80
	SD	1.83	1.94	2.01
Total	*M*	18.75	25.47	21.20
	SD	4.19	7.85	2.93
	n	16	15	15

elaborate on how spatial displays can be best used by teachers in their classrooms. First, concerning test validation, it might be recommended that teachers take into consideration that what is tested should indeed be part of the presentation of new material(s). In other words, by having special adjunct displays either (a) prepared by the teacher or (b) constructed by the students, the teacher should be cognizant of what will actually be tested. Therefore, both teacher and student benefit from this advance organization. Second, these studies suggest that there are definite biases between adjunct display construction and test construction when the constructor is one in the same. That is not a problem. The problem arises when the person who constructs the adjunct display(s) is different from the person constructing the test(s). In other words, what ecological validity do third-person adjunct display constructors have? Much more needs to be done in terms of the research attempting to answer this question. The authors agree that there is much variation between the two. Third, when the adjunct display constructors are the same as the test constructor, the effect of adjunct displays on transfer of information on application tests are much higher than with text alone. This finding is consistent with the previous literature. Finally, the authors would like to recommend that teachers use precaution when using a third-party test constructor or adjunct display constructor when they are not one in the same. Furthermore, it is recommended that it may be best to have students construct their own adjunct displays to accompany the traditional expository text.

References

Ausubel, D. P. (1968). *Educational Psychology: A Cognitive View*. New York: Hold, Rinehart, and Winston.

Barron, R. F. (1969). The use of vocabulary as an advance organizer. In H.L. Herber & R.T. Vacca (Eds.), *Research in reading in the content areas: First year report*. Syracuse, NY: Syracuse University Reading and Language Arts Center.

Davidoff, L. L. (1976). *Introduction to Psychology*. New York: McGraw-Hill, pp. 398-418.

Dembo, M. H. (1994). *Applying Educational Psychology*. White Plains, NY: Longman Publishing Group.

Earle, R. A. (1969). Use of structured overview in mathematics classes. In H.L. Herber & P. L. Sanders (Eds.), *Research in reading in the content areas: First-year report*. Syracuse, NY: Syracuse University Press.

Jonassen, D.H., & Marra, R.M. (1994). Concept mapping and other formalisms as Mindtools for representing knowledge. *Association for Learning Technology Journal, 2(1)*, 50-56.

Katayama, A. D. & Robinson, D. H. (2000). Getting students partially involved in note taking using graphic organizers. *Journal of Experimental Education, 68*, 119-133.

Kiewra, K. A., & DuBois, N. F. (1998). *Learning to learn: Making the transition from student to life-long learner*. Boston: Allyn & Bacon.

Levin, J. R., Bender, B. G., & Pressley, M. (1979). Pictures, imagery, and children's recall of central versusperipheral sentence information. *Educational Communication & Technology, 27*, 89-95.

Levin, J. R., & Berry, J. K. (1980). Children's learning of all the news that's fit to picture. *Educational Communication and Technology, 28*, 177-185.

Novak, J. D. (1990). Concept mapping: A useful tool for science education. *Journal of Research in Science Teaching, 27*, 947-949.

Robinson, D. H., & Kiewra, K. A. (1995). Visual argument: Graphic organizers are superior to outlines in improving learning from text. *Journal of Educational Psychology, 87*, 455-467.

Seifert, K. L. (1999). *Constructing a psychology of teaching and learning*. Boston: Houghton Mifflin Company.

Rubin, Z., & McNeil, E. B. (1981). *The Psychology of Being Human*. New York: Harper & Row.

Woolfolk, A. E. (2000). *Educational psychology (eighth edition)*. Boston: Allyn & Bacon.

Author Note

The authors would like to thank William Winn, professor of education at the University of Washington for his participation as a discussant and Bonnie Armbruster at the University of Illinois for her assistance in constructing the tests for this research that was presented as a symposium at the annual meeting of the American Educational Research Association.

Andrew D. Katayama is an assistant professor of Educational Psychology, Research, and Statistics in the College of Human Resources and Education at West Virginia University.

Daniel H. Robinson is an assistant professor of Educational Psychology at the University of Texas at Austin.

Kenneth A. Kiewra is a professor of Educational Psychology at the University of Nebraska-Lincoln.

Nelson Dubois has recently retired from the State University of New York-Oneonta where he was professor and department chair of Educational Psychology.

David Jonassen is a distinguished professor of Information Science and Learning Technologies at the University of Missouri-Columbia.