HUMAN-COMPUTER INTERACTION, 1998, Volume 13, pp. 1-35

Promoting Active Learning: The Role of System Structure in Learning From Hypertext

Amy M. Shapiro
University of Massachusetts, Dartmouth

ABSTRACT

Since its conception, hypertext has attracted the attention of educators and psychologists alike. Although a great deal is known about learning from text, little is understood about the process of learning from hypertext or what benefit it offers over traditional text. This study is an attempt to (a) determine whether theories of learning from text may be extended to learning from hypertext, and (b) learn more about the general effectiveness of hypertext on learning. Seventy-two college undergraduates participated in a study of hypertext-based learning. Each participant was assigned to work with one of three hypertext systems. All systems contained the same documents. Two of these contained the same electronic links (pathways) between documents and the third system condition served as a control. It was presented as a digitized book (linear text) rather than as a linked system. Participants' navigation behavior was logged electronically as they worked. Posttests included an essay, a series of short-answer questions, and a concept mapping task. Analyses revealed that learning from hypertext bears many similarities to learning from text, as the predictions made by Kintsch's (1988) construction integration model were borne out. System structure systematically altered what was learned from hypertext, just as characteristics of text alter text-based learning. System structure was also relevant to the way in which learners approached the material, as navigation behavior was affected. The less structured system seems to have promoted more active processing and a deeper

Amy Shapiro is a cognitive psychologist with an interest in human memory and learning; she is an Assistant Professor in the Psychology Department of the University of Massachusetts at Dartmouth.

CONTENTS

level of learning. In addition, hypertext was revealed to have only limited educational benefit for users in this study. Although no benefit of either hypertext system was observed over the linear system on the essay or short-answer questions, it was revealed that the presence of system links affects internal representations: Participants who were exposed to the hypertext systems produced concept maps that largely reflected their system links. Results are examined with reference to cognitive theory and the implication for system design is discussed.

1. INTRODUCTION

The introduction of computers and software into classrooms has attracted much attention from researchers and instructors over the past decade. One advance that has received a great deal of consideration in journals, conferences, and trade publications is hypertext.[1] Because this

1. *Briefly explained, hypertext allows large collections of text, graphics, sound* resources, demonstrations, and video to be displayed on small, single computers or as part of networked systems. Perhaps the most striking aspect of the technology

technology is a relatively new arrival in educational settings, relatively little is known about how students learn from it. Certainly, accepted guidelines for optimal system design for learning are yet to be established. It is the case, however, that theorists have spent a great deal of time examining the interaction between human information processing and traditional text. The effect of text structure on learning has been particularly well studied and used to inform theories about how we engage in this type of learning. The primary purpose of this investigation was to determine whether theories of learning from text may be extended to learning from hypertext. More specifically, its goal was to determine whether the global structure of a hypertext system effects learning in the same way as the overall structure of a piece of text. A secondary purpose was to learn more about the general effectiveness of hypertext on learning.

These goals are complementary because the structure of more traditional instructional material (e.g., text and lectures) has been shown to be relevant to the learning outcome (Bower, Clark, Lesgold, & Winzenz, 1969; Eylon & Reif, 1984; Kintsch, 1974). Until more is known about learning from hypertext, it is prudent to examine the effectiveness of the technology on more than one system structure so that we may separate the effect of any particular system's architecture from the state of the technology in general. In turn, testing a variety of system structures will allow the learning outcome of individuals working with differently structured systems to be compared so that the role of information structures in learning may be better understood. Before describing the study, the following section reviews some cognitive theory that may bear on learning from hypertext. The subsequent section reviews studies that have already contributed to our knowledge of the topic.

1.1. Cognitive Theory

For the purposes of this inquiry, a single document in a hypertext system is very similar to a piece of traditional text. Both media provide prose and perhaps illustrations. Hypertext may also contain sound resources or demonstrations. However, because the research reported here is concerned with the effect of system structure, these resources are not relevant here. In short, there is no reason to expect that reading single screens of text in hypertext systems should involve different processes: The relevant aspects of reading and comprehension are the same regardless of whether the print appears on paper or a screen. As Landow (1992) put it, "We still read according to print technology" (p. 41). The important

---

is the flexibility it offers programmers and users to interconnect or *link* conceptually related pieces of information with electronic "buttons" that move the user from one document to another when "clicked" with a mouse.

difference between hypertext-based information and traditional text-based materials is in the network of links connecting nodes. This network provides a richly integrated web of material. System interface and link identities may be sculpted to support an endless possibility of global structures. The relevant question is whether these links offer anything of value to the learner and, if so, whether the particular structure of the system alters the learning experience.

There has been insufficient time for the field to formulate a detailed theory of learning from such systems. However, theories of text-based learning provide a head start in that direction. Van Dijk (1980) proposed that learned information is incorporated or stored in *macrostructures*. His theory suggests that *information units* are linked together to form these macrostructures that serve to both organize and reduce complex information. Van Dijk argued that macrostructures allow us to form larger "chunks" of information that "have their proper meaning and function" (p. 14). Further, he contended that assigning a macrostructural unit to a series of independent facts (that he called "units at the microstructural level" [p. 14]) defines the thematic relation between them.

One may predict, then, that information structure may serve as an aid to forming these internal structures. Several studies exploring the effect of text structure on learning have shown that structures such as hierarchies can augment learning. For example, Eylon and Reif (1984) gave two groups of individuals the same body of information about a physics problem. One group was given a hierarchical organization and the other a linear organization. The researchers found that individuals in the hierarchical condition were better able to solve problems related to the materials they studied than those in the linear condition. In a study of list learning, Bower et al. (1969) also found that groups of words are two to three times better remembered when organized hierarchically rather than in randomly ordered lists. Highlighting the hierarchical structure of a body of information may serve to define the thematic relations referred to by van Dijk (1980) so that the material may be more accurately incorporated into a macrostructure.

Kintsch (1988) extended van Dijk's theory with his construction integration model. He proposed that such macrostructures are formed in memory when information contained in a text is stored. These structures, which he called *text bases,* mirror the text's organization and are constructed from its semantic content. Another kind of structure, called a *situation model,* may also be formed during reading. Situation models contain the information from the text base and additional information from permanent memory. The situation model, then, may be thought of as the storehouse for our deeper understanding of written material. A more involved form of processing is required to create a situation model, as the new information must be integrated with prior experiences and knowledge. This idea has been a component of other theories of learning and memory. For example, the levels of processing view (Craik & Lockhart, 1972) also proposed that more active processing would lead to an increased capacity for learning.

According to the construction integration model, then, a text structure that provides the opportunity to incorporate knowledge within a text with our prior knowledge should result in superior learning. This prediction was borne out in a study by McNamara, Kintsch, Songer, and Kintsch (1996) when they provided expert and novice students of biology with texts of either high or low coherence. Although the texts used in the study had approximately the same content, the difference between versions was that "the interconnections between the various content units [were] made explicit" (p. 8) in the high-coherence version. They found that the experts learned more from the low-coherence text whereas the novices learned more from the high-coherence text. McNamara et al. claimed that the experts had sufficient prior knowledge to incorporate with the text and form a situation model when working with the low-coherence text. The high-coherence text was not challenging enough and provided no opportunity to use prior knowledge. The novices, on the other hand, had insufficient knowledge to draw on, and so were unable to move beyond the creation of a text base. For this reason, they did better with a more coherent text that was more within their grasp.

For this study, the construction integration model would predict that, given equivalent text, reading single screens of text in hypertext systems should not result in different learning than reading from traditional text, as the relevant processes should be the same. The construction of text bases for factual information contained within the text of both media should be identical between traditional text and hypertext. It would also predict that learning from a hypertext system should promote learning of surface information regardless of its link structure.

Less certain is how the links in a network of pages or screens are incorporated into memory. It is likely that learners will be able to incorporate knowledge of system links into their understanding, just as they should with other factual information presented within a document of a linked system. Obviously, this would be an advantage over traditional text, which cannot provide system links. Will learners attain a deep understanding of the conceptual meaning of those links (i.e., create a situation model for this information)? The construction integration model would predict the development of a situation model if the learner engages in the necessary processing to incorporate prior knowledge with the information in the system. Because the results of McNamara et al.'s (1996) study showed that learners attain a better understanding of text when required to put in more thought, a hypertext user should perform better when put into a position to consider more deeply the conceptual meaning behind a link between documents. Given this, the construction integration model predicts differences in learning due to system structure if the structures differ in the way in which they promote thinking about the systems' conceptual links.

1.2. What Do We Know About Hypertext-Assisted Learning?

The previous section highlights the importance of active learning in practice as well as in theory. Investigators and educators have had high hopes for the effectiveness of hypertext as an educational tool because of its potential to facilitate active learning. Leggett, Chase, and Kacmar (1990) labeled the potential of hypertext to engage the reader as an active learner as the "revolutionary" aspect of hypertext. Landow (1992) suggested that students using hypertext systems are active learners in part by choosing which links to follow. He quoted Jonassen and Grabinger, who stated that "hypermedia users *must* be mentally active while interacting with the information" (1990, quoted in Landow, 1992, p. 121). If this statement is true and the construction integration model's prediction about the effect of active learning is also accurate, the use of hypertext should prove to be an effective way of acquiring a deep, integrated understanding of a domain.

It is very possible, however, to use hypertext in a passive way. Indeed, Meyrowitz (1986) noted that some users of the Intermedia hypertext system showed little or no benefits from its use. He suggested the reason for this failure in these users was that they did not engage in active use of the system. Hammond (1990) recognized this problem and suggested that educators take some of the control from users by providing tasks or assessments that would require students to approach the material in a more active manner.

Hammond's (1990) suggestion would appear to be appropriate because the very ability of the user to be in control of the learning situation may contribute to difficulties in learning from hypertext. For example, in a study of mathematics learning, McGrath (1992) investigated the effects of user control and spatial ability on how students interact with hypertext and what they learn from it. She tested students who worked with either standard hypertext, a traditional text-based lesson on paper, or one of two computer-assisted instruction (CAI) conditions. The hypertext group had the most control over their learning and the CAI groups had the least. Results showed that the outcome of the sessions depended on the system as well as the learners' spatial ability. There was an overall effect of system condition on near transfer problem solving, with the participants in the paper-based group performing best and the CAI participants performing most poorly. Although there was no significant interaction between spatial ability and system condition, McGrath reported that the highest scoring group was the high spatial/paper group and the lowest scoring group was the low spatial/hypertext group. These results were explained by McGrath by looking to the learner control literature, which points to the difficulty of low-ability learners in undertaking tasks with a high degree of control. These results point to possible drawbacks of hypertext use by low-ability learners. They also put into doubt whether hypertext offers advantages over text-based learning.

In spite of such discouraging results, other studies point to the benefit of hypertext on learning. Beeman et al. (1987) reported that they incorporated the Intermedia hypertext system into a biology course and an English literature course. Their aim was to observe the progress of the students in these courses and to compare their performance in their respective classes with that of students from previous semesters who had taken the same courses without access to Intermedia. Beeman et al. reported that the English professor felt that class discussion had improved. In addition, they felt that the quality of the essays written by these students was markedly superior to that of essays written by students in previous semesters. These conclusions, however, are fairly subjective and there is no way to be certain whether the data in question reflect measures of deep or surface learning. Beeman et al. did report, though, that students in the biology course performed better than those from previous semesters: 44% of the Intermedia students received a grade of A in the course, whereas only 34% had received a grade of A in the previous semester. However, Beeman et al. did not report whether this difference was significant and there is no indication that the respective academic abilities of the test and control classes were factored into their assessment. In sum, although the results of this study offer encouraging news about the effects of educational hypertext, they do not indicate what the underlying source of the results may have been.

Other evidence has come out of the ACCESS Project. This program, operating in several high schools in Providence, Rhode Island, and Eugene, Oregon, involved the use of a large corpus of material on American art, history, culture, and literature from 1607 to 1970. Spoehr and Shapiro (1991) observed the progress of students who used ACCESS over the course of a year in an Advanced Placement history course. They compared the Advanced Placement exam scores in history between the ACCESS students and those who had taken the same course with the same instructor but without ACCESS the year before. Spoehr and Shapiro reported that the ACCESS students performed significantly better on the exam than the non-ACCESS students. The scores rose from a mean of 3.22 to a mean of 3.81 (out of a possible score of 5.00). Even after conducting stepwise multiple regression analyses to partial out the effects of general academic ability, the results suggest that using ACCESS had a positive effect on how well students were able to master historical material and perform on a standardized test. Again, although such results speak positively for the effect of hypertext in educational settings, the question of exactly what was learned (the text base or situation model) by students to augment Advanced Placement exam performance is unanswered.

The results of various studies discussed here vary quite a bit in terms of their conclusions about hypertext learning outcomes. This seeming inconsistency is echoed in a meta-analysis of experimental hypertext studies reported by Chen and Rada (1996). Among many other analyses, the authors looked at 13 studies that compared the effectiveness of people using hypertext systems with that of people using nonhypertext systems. Eight of these studies reported higher effectiveness of the hypertext group; the remaining five studies reported the opposite results. Chen and Rada suggested that underlying factors such as differences in system design, different substantive material in hypertext documents, and the design of the experiments might be the cause of these differences. Understanding the root of these differences is important because such knowledge would be an indication of the reason underlying any benefit to the learner and the processes involved in attaining such benefits.

Because the construction integration model predicts that information structure could effect learning outcome by altering the way students work with the material, this study focuses on system design as a possible source of hypertext's effectiveness, as suggested by Chen and Rada (1996). Three hypertext systems of varying structure were created for the study. Students were each asked to work with one of these systems so that their learning outcomes could be compared.

2. A STUDY OF LEARNING FROM HYPERTEXT

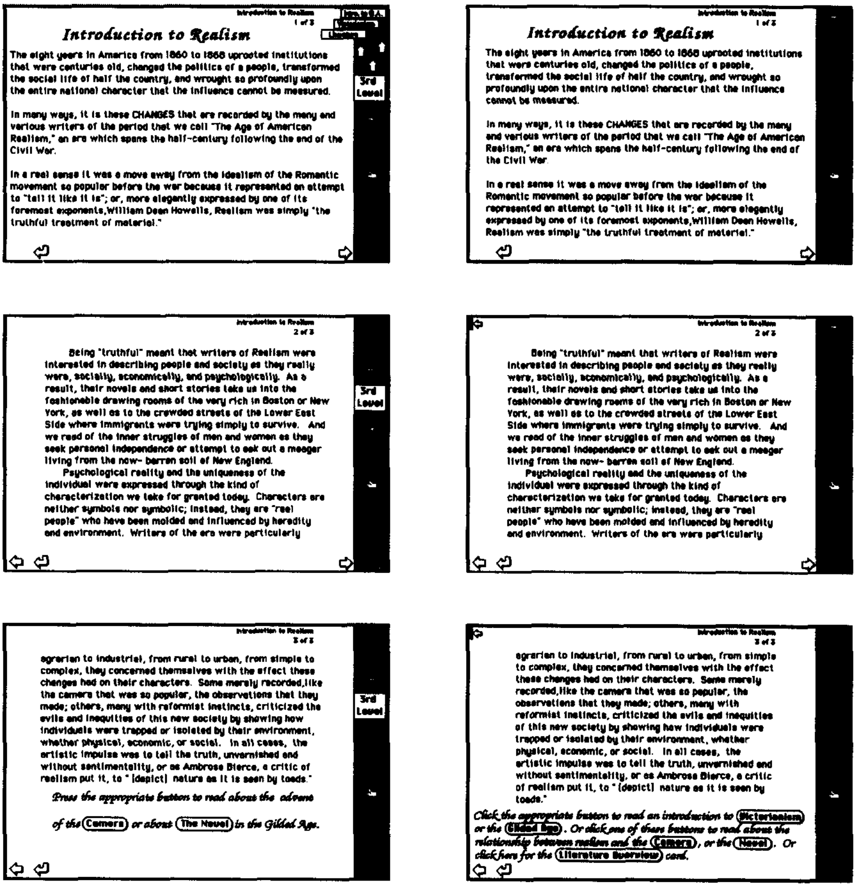
Participants in this study were all pretested for knowledge of history and assigned to one of the three systems to work with during the subsequent learning phase of the study. They were each then given a task to complete as their goal during the learning phase. Once this task was completed, all participants were given a variety of posttests. The following sections describe the systems, goals, and testing materials in detail.

2.1. System Conditions

The hypertext materials used in the learning phase of the experiment were originally created by several high school teachers from Providence, Rhode Island. Some of the materials were authored by the teachers, and some were borrowed from books, articles, and other sources. For the purposes of this study, the teachers' materials were redesigned to create a HyperCard corpus of 46 documents. Documents are made from *cards,*

***Figure 1.* A typical document in the highly structured (HS) and unstructured (US) system conditions.**

Highly Structured Condition | Unstructured Condition



independent screens of information. Each document in the systems created for this study was between one and five cards long. The left- and right-hand panels of Figure 1 each provide an example of a single document composed of three cards. The material covered a range of topics—such as art, history, culture, and literature—about the "Gilded Age" in the United States, the period spanning from 1865 to 1900.

The document content of the systems used in this study was identical across conditions, regardless of the way the system was organized. In all three systems, participants were able to page backward and forward within a document by using the Previous and Next buttons contained on each card.[2] (The first card of a document, obviously, had no Previous button, and the last card had no Next button.) Two of the systems were constructed as multiply linked hypertext systems and contained between one and five links to other documents (nodes), all of which were identical between these conditions. The third (control) condition, by contrast, was not designed as a multiply linked hypertext system, but a digitized book. Each document in that system was linked to only two documents: that which preceded it and that which followed it in the linear sequence. (By default, of course, the first and last documents were each only connected to a single document.)

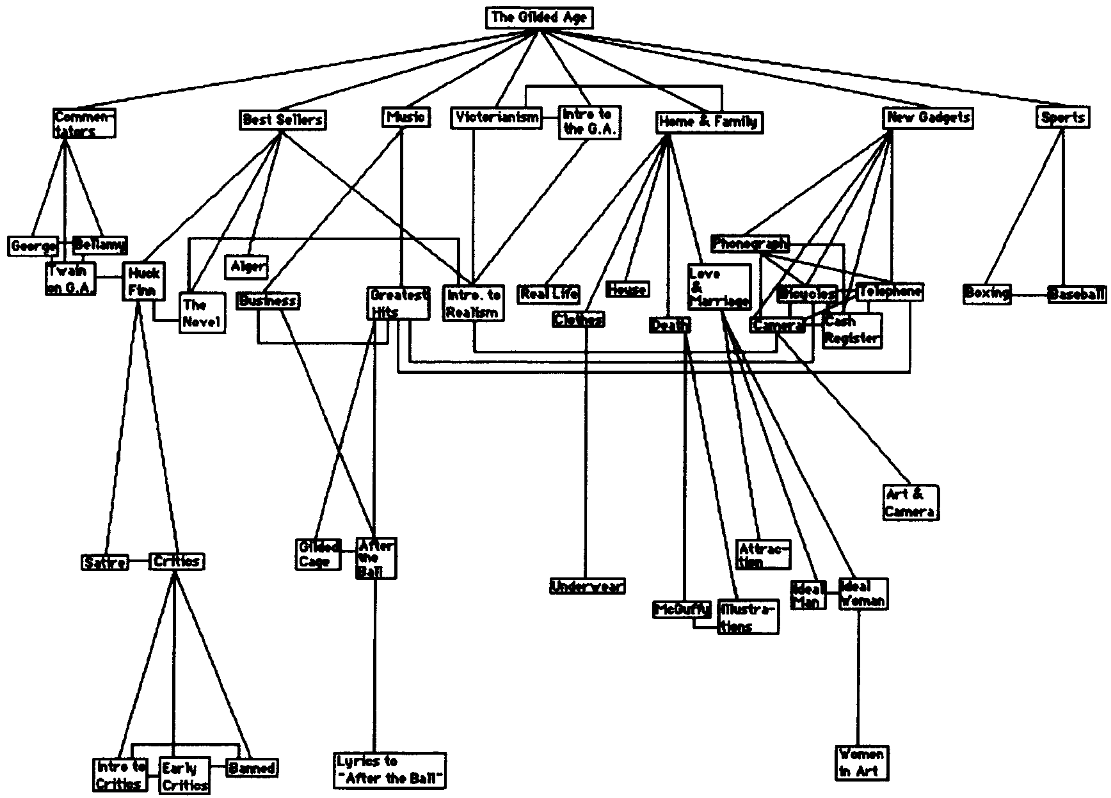
One of the experimental systems was hierarchically structured. The purpose of including this structure was to expose participants to a highly structured system that provided information about the nature of the relations between documents. In addition, the McNamara et al. (1996) study points to the benefit of a highly coherent text for low-domain-knowledge learners and all participants in this study had low domain knowledge. In addition, research discussed earlier has pointed to the benefit of hierarchical structures in particular with reference to text-based learning. This structure was chosen, then, to test the effects of a highly organized system on hypertext learning. The other system was designed to appear to subjects as an unstructured system of nodes and links. It provided the same links as the highly structured system but offered no overt clues as to their meaning. Because the system had no obvious overall structure, participants interested in understanding the relations between linked documents would have to engage in the processing necessary to come to an understanding on their own. These two conditions, then, were included to evaluate the formation of situation models, which the construction integration model predicts are established when active processing takes place. The linear system served as a control that allowed comparisons between these systems and a (more traditional) linear presentation of text. This condition permitted exploration of the second goal of this study, to learn more about the general effectiveness of hypertext as a learning tool. Detailed information about each of these systems is provided in what follows.

The *highly structured (H)* system was structured so that very general information about the Gilded Age was presented on a single document, located on the first level. The second level presented major subtopics covered in the system, and subsequent levels presented further subtopics or provided more specific information. This organization is illustrated in Figure 2.

---

2. Two other conditions were also developed in order to explore issues related to this investigation. Neither these conditions nor the data they provided are reported here.

Figure 2. A pictorial representation of the organization of the highly structured condition.



The system was designed to appear to users as a hierarchically structured network of nodes and links. Their movement between documents was either lateral (across a single level) or vertical (up or down a branch to a new level). Navigation tools were provided to encourage the feel of moving through the branches and levels of the hierarchy. As illustrated in the three left-hand panels of Figure 1, for instance, each page in Condition H had a field in the right-hand margin that contained the level of that particular document. So, for example, each page of the document about the Realism movement contained a field marked *3rd Level,* as that document was on the third level of the hierarchy.

The buttons that linked the current document with its superordinate documents appeared as arrows that pointed from the current level field to the superordinate card name field. This field listed the name(s) of the document(s) that was (were) above the current document—its *superordinate* document(s). In the example provided in Figure 1, the first page of the *Introduction to Realism* document had fields that read Literature, Victorianism, and Introduction to the Gilded Age above the current level field. This was done to help orient users to where they were in the stack's structure. The use of the arrow icon pointing up to the superordinate topic names was intended as a reminder that clicking one of these link buttons meant moving up in the hierarchy. Links to lateral (same-level) or subordinate documents were generally located on the last card of a document.

Visual effects were also used to encourage the impression of moving between or within levels. For example, when a button was clicked to move the user vertically to a superordinate document, the screen did not simply flash to the new document. Rather, the new card slowly replaced the old from the top of the screen to the bottom. This visual effect is referred to as *wiping down.* Likewise, when moving laterally, the screen slowly wiped left.

Other tools were used to help orient participants. The user often needs to know where he or she has been in order to get situated in the learning space (Hammond & Allinson, 1989; Landow, 1992; Nielsen, 1990). In fact, Hammond (1990) made the point that without facilities to provide "footprint" information (where have I been so far), users may have difficulty using a system. Therefore, a tool called a *marker,* represented by the "string finger" icon, was available on every card. When the marker icon was clicked, a field appeared that provided the name and level of the document the user visited before arriving at the current document. The user was given the option of either returning to the previous document or of resuming the learning session as though it had never been interrupted.

There was also a tool called the Hansel & Gretle button that allowed users to backtrack along their path, as if picking up bread crumbs they dropped as they moved through the system. This was represented by a curved arrow on the lower, left-hand portion of every card. If a user had visited four cards, for example, clicking this icon three times would take him or her from the fourth card to the third, then the second, and finally to the card of origin. The purpose of this tool was to keep participants from getting lost, which is a common problem in hypertext learning (Nielsen, 1989, 1990).

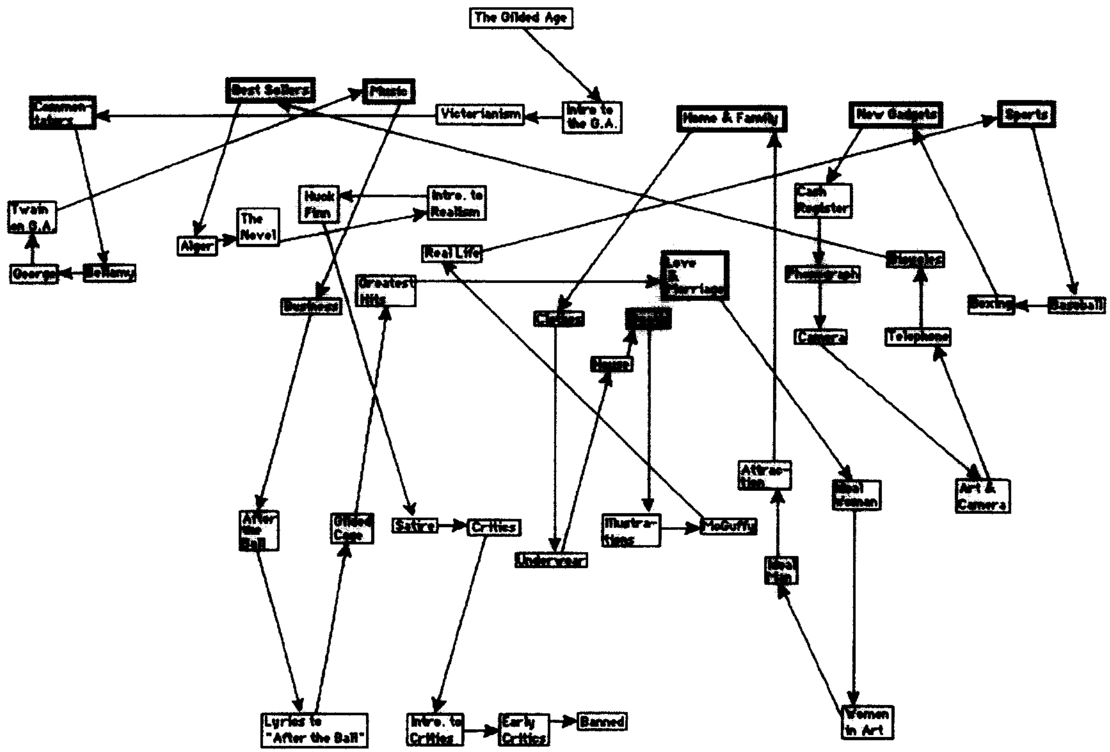
The *unstructured (U)* system provided the same links and documents as the highly structured system, but provided the user with no explicit information about the system's macrostructure. Participants working with this system navigated through the material using link buttons, just as those in the highly structured system did, although there were no orienting tools to provide clues about link relations and system structure. All links simply appeared at the end of each document. However, this system also contained the Hansel and Gretle button. Sample documents from this system are also provided in Figure 1.

The *linear (L)* condition was a control. There was no hierarchy and no interconnectivity between documents—except, by default, those that happened to be next to one another in the series of cards. It was presented as a book, divided into chapters that were defined by topic (e.g., music, commentators, sports, etc.). Each chapter heading, however, corresponded to a major topic in the other two conditions. These topics roughly correspond to the topics on the second level of the highly structured system's hierarchy. Figure 3 uses the highly structured map to illustrate the path taken by participants in the linear condition (which started with the Gilded Age overview card) as they worked through the system. Nodes that are boxed by dark borders were identified as chapter headings.

2.2. Task Assignments

It is well understood that students' goals have great impact on the way in which they approach the tasks of learning, studying, and remembering (Edwards & Hardman, 1989; Hammond & Allinson, 1989; Hulse, Deese, & Egeth, 1975; Leonard & Whitten, 1983; Light & Carter-Sobell, 1970; Marchionini, 1989). We process or encode information differently, depending on how we believe we will eventually be called on to use it—task demands affect the strategies we use to learn and remember information (Hulse et al., 1975; Leonard & Whitten, 1983; Light & Carter-Sobell, 1970; Murphy & Shapiro, 1994). Because student goals are known to affect the learning outcome, asking participants to engage in any single learning task could be a confounding factor. That is, any given goal could alter what is learned from one system design but not another. For this reason, task was used as a control variable. Participants were given one of three goals prior to the learning phase of the experiment. They were either asked to (a) answer an integrative essay question; (b) answer a series of short-answer questions, designed to probe for simple facts; or (c) browse through the system, reading what is found to be of interest. These groups are referred

Figure 3. **The path followed by participants in the linear condition, illustrated on the highly structured map.**



to as the *integrating, fact-finding,* and *browsing* groups, respectively. These goals were chosen because studying for the purpose of integrating and understanding information from a variety of sources and finding independent, specific facts are two common tasks faced by history students. In addition, students are often given reading assignments without being offered a specific goal. The three goal conditions were created to approximate these types of classroom or homework assignments. The integration and fact-finding tasks offer additional benefit because they will put individuals to the explicit task of either amassing facts or integrating information, which are the outcomes of interest here.

2.3. Participant Pool and Pretesting

Participants were recruited from the Brown University undergraduate community. Only native English speakers between 18 and 30 years of age with no known learning disabilities were admitted into the study. Each eligible participant was paid $12 for his or her participation in the study, which lasted an average of roughly 2 hr. Variability in time among participants was dependent on the individual's work pace.

All potential participants were pretested for their knowledge of American art, culture, history, politics, and literature of the latter half of the 19th century, the period known as the Gilded Age. The 20 items that comprised the pretest were short-answer questions that probed for knowledge of very basic facts about the Gilded Age in the United States. They were designed by the experimenter to probe general information found in a high school textbook and approved by a high school history teacher prior to use. The purpose of the pretest was to ensure that participants in the study had little or no knowledge of the topic they were to study during the learning phase of the experiment. Because the system dealt with very specific facts about that era in history (as consequently did the posttests) and the pretest questions addressed general issues, it is unlikely that an individual scoring low on the pretest would know much about the contents of the corpus and do well on the posttest. Therefore, only those scoring below 45% correct were included in the study. A test of the pretest materials, which is not reported here, was conducted to determine the acceptable range of pretest scores.

In all, 144 individuals took the pretest. Seventy-two individuals participated in the study, 13 scored too high on the pretest to qualify as participants, 11 were unable to schedule a convenient time to participate in the experiment, and 48 were randomly assigned to participate in conditions for a related study. Qualified individuals were scheduled to participate in the study anywhere from 1 day to 2 weeks after taking the pretest, depending on their schedules.

2.4. The Learning Phase, Posttesting, and Dependent Measures

The study reported here is a completely randomized 3 × 3 factorial design. Participants were assigned to one of three task conditions and to one of three system conditions. Eight participants were run in each of the study's nine conditions.[3] Each was instructed about how to use the system and each worked on a Macintosh IIsi. Participants were told how to click on buttons that would allow them to navigate between documents. They were told to ask any questions they had at any time during the session, to take as long as they needed to complete their research, and that they would be given various tasks to perform after they completed their work. Those in the linear condition were told to use as much time as they needed, but that they should read to the end of the materials. All participants were told that they would be tested on what they had learned at the end of the learning phase. Those who were assigned to the integrating or fact-finding conditions found their questions attached to the assignment sheet.

Participants in the browsing condition were told to go through the system and learn as much as they could about its content. They were told that they would be tested on its content when they were through.

Participants in the fact-finding group received 10 questions to research, the answers to which were located on separate system documents. Examples of short-answer questions include the following:

- How was the success of a song measured for its writer and publisher in the Gilded Age?
- In what significant way did Horatio Alger's novels depart from their otherwise traditional Protestant tone?
- What was the "ideal woman" during the Gilded Age?
- In what ways did the Victorian obsession with death change the funeral?

The essay question previewed by the integrating participants was the following:

The Gilded Age was a time of great change. This made it a paradoxical era and one in which there was great tension between old tradition and modern thinking. Using as many aspects of American life and culture during that time as possible, explain why this is so.

---

3. As stated earlier, the purpose of including multiple task conditions was to avoid possible confounds from any single task assignment. As reported later, there were no interactions between goal condition and system condition so the system structure analyses were all collapsed over goal treatments to create a total of 24 participants per system condition.

Participants were allowed to navigate freely through the system and to take as long as they needed to complete their research. As they worked, an internal program recorded their navigation behavior. This program provided information about each link traversed by participants, how long they stayed on any particular document, the nature of their search patterns, which navigation tools they used, and so forth. As it is likely that participants' study behavior would be related to posttest performance, it was important to learn about the effect of system structure on navigation. For this reason, navigation behavior was treated as a dependent variable.

When participants were through studying, the computers were turned off and the testing phase began. Regardless of which task condition participants were assigned, all were asked to write an essay on the same topic that the essay participants received in advance, answer 30 short-answer questions (all of which were novel to all but the fact-finding group who had received 10 of the 30 questions for their assignment), and draw a concept map of the topics covered in the system. The short-answer posttest was administered to obtain a measure of factual knowledge acquired by participants (the text base). The essay posttest was offered as a measure of the depth and integration of individuals' understanding of the instructional materials (the situation model). The concept mapping task was administered to obtain some measure of participants' conceptual organization, including how much they learned about topic relations from the materials. Participants were given the posttests in a randomized order and allowed to take as long as they needed to complete them.

3. RESULTS

As explained earlier, multiple goal conditions were used to avoid possible confounds of any single task assignment. No significant interactions were found between system and task conditions for any of the dependent measures (navigation behavior, essay performance, short-answer score, and concept mapping results), so the analyses reported here for each dependent measure were all collapsed over task condition, resulting in a total of 24 participants per condition.

3.1. Integration of Knowledge

The posttests discussed in this section were aimed at testing the degree to which learners were able to integrate the information presented by the system with their prior knowledge. Rather than testing for factual knowledge, these tests assessed the information that learners were able to infer or create on their own. These tests, then, were all measures of what Kintsch (1988) called the situation model. The prediction made for this study was that the unstructured system would show greater benefit to learners because it more readily lends itself to the kind of processing necessary to create an integrated understanding of the material (e.g., active learning).

Essay Posttest

A content expert, unaware of the aims, methods, or treatment conditions of this study, was hired to assess the essays on a variety of dimensions related to the participants' depth of understanding of the material. The expert was a doctoral student in the Department of History at the University of California at Berkeley who specializes in 20th-century U.S. history. She is the author of several conference presentations and published papers and has graded hundreds of college-level essays as a teaching assistant for the university. Four dimensions of the essays were chosen for their relevance to the stated goals of this investigation:

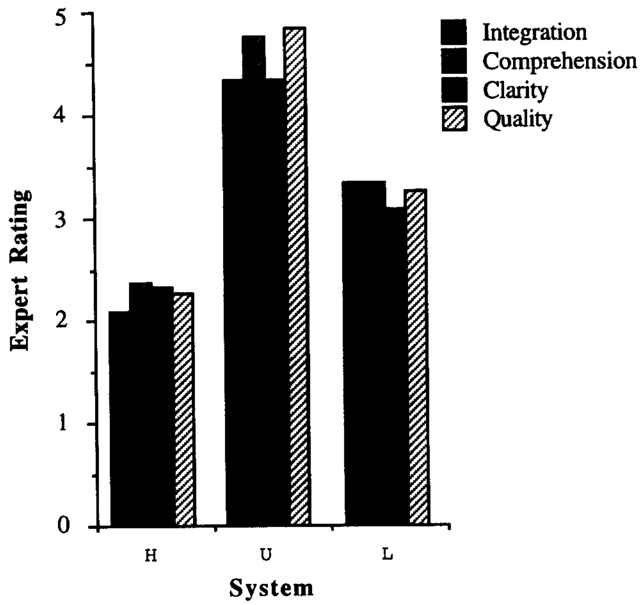
1. How well integrated was the information in the essay?
2. How clear was the author's argument?
3. How deeply does the author understand the topic about which he or she is writing?
4. How was the overall quality of the essay?

Before the content expert rated the essays, the reliability of her ratings was established. Fifteen essays were chosen at random and both the expert and the experimenter,[4] blind to system condition, rated them using a 7-point Likert scale (the higher score indicating a more positive rating) on the dimensions just described. The ratings of each judge for the four dimensions were found to be highly correlated. For Dimensions 1 through 4, the correlations were $r(13) = .67$, $p < .01$; $r(13) = .68$, $p < .01$; $r(13) = .81$, $p < .01$; $r(13) = .76$, $p < .01$, respectively.

Once reliability was established, 36 essays representing an equal number from each task and system condition (for a total of 12 essays from each system condition, divided evenly among the goal groups) were chosen at random and rated by the content expert. As illustrated in Figure 4, all analyses were significant. There was a reliable effect of system structure on how well integrated the essays were, $F(2, 33) = 3.41$, $p < .05$. Post-hoc analyses using Scheffé's F test show that the highly structured group's mean rating was significantly lower than the unstructured group's, $F_{\text{comp}} = 3.41$, $p < .05$, but not the linear group's.

4. Because I worked intimately with the system materials while creating the stimuli, have a degree in French literature, and worked for many years as a writing tutor, I felt it was appropriate to serve in this capacity.

Figure 4. **Mean ratings derived from the content experts' evaluations of the essay posttests.**



The depth of participants' understanding of the material also differed between system conditions, $F(2, 33) = 6.00$, $p < .01$. The highly structured group's mean rating was significantly lower than the unstructured group's, $F_{comp} = 5.822$, $p < .05$, but not the linear group's.

The clarity of participants' essays also varied between system conditions, $F(2, 33) = 4.48$, $p < .01$. Again, the highly structured group's mean score was reliably lower than that of the unstructured group's, $F_{comp} = 4.39$, $p < .05$, but not the linear group's.

Finally, the overall quality of participants' essays differed between system conditions, $F(2, 33) = 6.32$, $p < .01$. The highly structured group's mean score was reliably lower than that of the unstructured group, $F_{comp} = 6.06$, $p < .05$, but not the linear group.

In short, an interesting pattern emerged from the essay data. On every measure of essay quality related to depth of content, the highly structured participants performed more poorly than their counterparts in the unstructured group. The performance of the linear group was statistically comparable to that of both of the experimental groups.

Concept Map Posttest

To assess the complexity of participants' conceptualization of the material, a measure of link density for the concept maps was derived by dividing the total number of links in each concept map by the total number of nodes. There was no significant effect of system structure on concept map link density, $F(2, 68) = .406$, $p > .05$. The highly structured, unstructured, and linear system groups had mean link density scores of 1.04, 1.01, and .97, respectively.

3.2. Factual Knowledge Acquisition

Results of the posttests reported in this section all provide information about the amount of factual knowledge acquired by participants during the learning phase of the study. All of the tests were aimed at assessing the factual knowledge presented within the system rather than information one would have to produce or infer in any way. These tests, then, were all measures of what Kintsch (1988) would call the acquired text base. The prediction made for this study was that there would be no effect of system structure on these measures.

Concept Map Posttest

There was a significant effect of corpus structure on the number of items participants included in their concept maps, $F(2, 68) = 13.12$, $p < .001$. The means for the highly stuctured, unstructured, and linear system conditions were 25.26, 24.79, and 17.5, respectively. Participants in the linear condition recalled fewer documents than those in the other conditions. Post-hoc analyses show that the differences between the linear group and the highly structured group, $F_{comp} = 10.29$, $p < .05$, and the unstructured group, $F_{comp} = 9.29$, $p < .05$, were significant.

The most likely cause of this result is the fact that participants in the two linked conditions saw the topic names many more times than those in the linear condition because the topic names were listed on the buttons used for navigation. It is possible, then, that reduced exposure to the topic names themselves is responsible for the difference in recall ability between the linear system group and the other two. A repetition effect is by no means extraordinary and given the nature of the data collected from this study, there is no way to show that the act of navigating links is responsible for this phenomenon. However, if presenting topic names a number of times is an effective way of augmenting memory for topics covered in a learning session, doing so through link buttons appears to be a workable approach.

A separate analysis was conducted to explore how closely the links drawn in participants' concept maps mirrored those in the systems. This was done to explore the effect of the links themselves on acquired knowledge of document relations. Toward this end, each link in participants' maps was categorized as either *familiar* (present in the two experimental systems) or *novel* (not represented in those systems). For example, a link between Realism and Victorianism would be categorized as familiar, because those documents were linked in both of the experimental systems. A link between Realism and Boxing, however, would be categorized as novel, because those documents were not directly connected in the experimental systems. Of course, there were no links between system documents in the linear condition, but judging that group's concept maps by this criterion provides a baseline measure for the other groups.

An analysis of variance revealed that there was a significant effect of system structure on the number of familiar links present in participants' concept maps, $F(2, 68) = 10.85$, $p < .001$. The mean number of familiar links for the highly structured, unstructured, and linear groups was 69.94, 73.76, and 48.52, respectively. The linear group had significantly fewer familiar links than the highly structured group, $F_{comp} = 6.61$, $p < .05$, and the unstructured group, $F_{comp} = 9.37$, $p < .05$. Participants in the linked conditions incorporated topic relations to which they were exposed into their own stored representations. However, differences between the two hypertext system structures did not affect this outcome of the learning process.

Short-Answer Posttest

Although the results approached significance, there was no main effect of system type on students' performance on the short-answer posttest, $F(2, 69) = 2.81$, $p = .067$. The mean scores for the highly structured, unstructured, and linear conditions, respectively, were 65.83, 55.42, and 79.58. Although there was no significant interaction between system condition and goal condition, the slightly better performance by the linear group may reflect the performance of the fact-finding group in this condition. In fact, when the fact-finding group is removed entirely from the analysis, the trend toward significance disappears, $F(2, 45) = .734$, $p = .49$.

Essay Posttest

A quantitative analysis was performed on the essays to assess the amount of factual knowledge that participants attained and were able to demonstrate in their essays. Toward this end, a coding procedure was developed. It was roughly modeled after Meyer's (1975) prose analysis procedure. The coding scheme involves representing each proposition, or unit of information in the essay, as a node. Each additional proposition that is related to it by the author is connected by way of a link. In this way, each essay is transformed into a series of nodes that are connected by links.

The final product of the procedure is an essay map for each essay, which may then be analyzed for, among other things, node density.

To ensure the reliable use of the prose analysis procedure, a measure of interrater reliability was obtained. Six essays were randomly chosen as samples. A second rater, who was blind to the design, hypotheses, and aims of this study, was trained on the analysis procedure. This rater and the experimenter independently mapped the six essays, and the results were correlated to obtain a reliability measure. Comparisons were made between the raters' judgments about node identity, defined as the inclusion of specific units of information in the essay.

A Spearman rank correlation shows that there was a significant correlation between the raters' judgments of node identity in the essays, $r = .5$, $p < .001$. It is worth noting that this measure was also significant between the two raters for each of the six independent essays. The correlations between essay numbers 1 through 6 were $r = .57$, $p < .01$; $r = .33$, $p < .05$, $r = .45$, $p < .05$; $r = .76$, $p < .05$; $r = .61$, $p < .01$; and $r = .58$, $p < .01$, respectively. Additionally, Pearson correlations were performed between the raters' judgments of the total number of propositions in each essay. This correlation was reliable as well, $r = .902$, $p < .05$.

Having established the reliability of the scoring procedure, the remaining essays were scored. An analysis of all participants' essays revealed no effect of system structure on the propositional content of participants' essays, $F(2, 69) = .263$, $p > .05$. The highly structured, unstructured, and linear proposition density means were 22.75, 22.96, and 21.21, respectively. It appears that all participants were able to include an equivalent amount of factual information in their essays, regardless of their respective system conditions.

3.3. User Behavior (Navigation Trails)

Because a secondary goal of this study was to learn more about how people learn from hypertext, the way in which participants interacted with the system is important. The navigation data revealed that the length of time participants spent reading through the system materials was significantly influenced by system condition, $F(2, 61) = 7.99$, $p < .001$. The means are provided in Figure 5. Post-hoc analyses reveal that the linear group spent significantly more time than the highly structured group, $F_{comp} = 5.82$, $p < .05$, and the unstructured group, $F_{comp} = 6.89$, $p < .05$.[5]

5. Eight of the linear participants' navigation data were lost due to system failure. Because that group's navigation was so constrained, however, relatively few relevant analyses stem from those data.

Figure 5. **Navigation and concept map means (and standard deviations) for the three system conditions including (a) length of total time spent researching, (b) time spent reading each document, and (c) percentage of documents read.**

	System Type					
	H		U		L	
Measure	M	SD	M	SD	M	SD
Total minutes	63.91	17.09	61.81	21.20	87.81	27.91
Minutes/document	1.64	.39	1.58	.62	1.91	.61
% read	87.82	6.96	87.23	6.62	99.75	0.14

As indicated by the means listed in Figure 5, there were no significant differences between system groups with respect to the mean length of time spent on each document, $F(2, 61) = 1.95$, $p > .05$. However, there were differences between groups in the number of documents they each read, $F(2, 61) = 17.35$, $p < .0001$. These means are also provided in Figure 5. Post-hoc analyses reveal that participants in the linear group read significantly more documents than those in the highly structured group, $F_{comp} = 15.08$, $p < .05$, and the unstructured group, $F_{comp} = 12.50$, $p < .05$.

This result may be due to the way in which users were forced to work through the material. Participants in the two linked conditions had the opportunity to skip over documents (by choosing not to move to a particular document during his or her travels), whereas those in the linear condition had no choice but to move through every document. However, participants in the linear system were still able to skip documents by clicking the Next Card button in rapid succession, thus leaving a document immediately after it appeared on the screen. The program that logged the navigation behavior of each participant was able to detect when participants displayed this behavior. Documents that were merely flipped through were not included in this analysis. Even with this conservative approach, however, the linear participants were shown to have read significantly more of the materials than those in the other conditions.

There was also a difference between participants in the two experimental conditions with respect to the way in which they navigated the corpus. In particular, the types of links students were inclined to follow varied with the appearance of the system. As emphasized earlier, participants in both experimental conditions were exposed to the same nodes and links, and all were able to navigate freely. However, each of these system organizations was designed to either emphasize or obscure the implicit hierarchical structure. The highly structured system was designed to enable participants to distinguish vertical from lateral links. *Vertical* links moved the user up or down a branch (between levels), whereas *lateral* links moved the user between documents on a single level. By contrast, all links appeared the same to participants in the unstructured conditions, who were not explicitly made aware of any hierarchical structure to the system. For the purpose of this analysis, each link was categorized as either lateral or vertical, although participants in the unstructured condition were not attuned to the distinction. The difference between groups' navigation patterns is made clear by analyzing the proportion of lateral to vertical links followed by participants in each of the system conditions. This measure is reported here in terms of the percentage of lateral moves out of total moves made.

A *t* test reveals that there was a significant difference between system condition groups' navigation patterns in terms of the relative number of lateral links they traversed, $t(45) = 3.86$, $p < .001$. The highly structured group made an average of 23.69 lateral moves whereas the unstructured group made an average of 15.42 lateral moves. This translates to 27.8% and 16.97% of total moves made by each group, respectively.[6] Implications of this result are discussed in the next section.

4. DISCUSSION

The primary goal of this study was to learn whether the structure of a hypertext system affects learning in the same way as a piece of text. This goal was met by testing whether the predictions made by theories of learning from text (such as the construction integration model) about learning from hypertext would be borne out. Results of this investigation show that, as predicted, system structure had a bearing on learners' ability to arrive at a deeper, more meaningful understanding of the information beyond that overtly stated in the material. However, it was also predicted that there would be no effect of system structure on the acquisition of general facts presented in the material. This prediction was borne out as well. A secondary goal was to learn more about the general effectiveness of hypertext on learning. The study revealed some limited advantage of hypertext on learning over that of a linear presentation. Each of these points is discussed in detail in what follows.

---

6. It is worth noting that even the hierarchical group's lateral link rate of 27.8% is much lower than the systems' lateral to vertical link ratio of 38.46%, so both system groups made many more vertical than lateral moves than might be expected by chance. This finding is most likely an artifact of the system's architecture. Figure 1 shows that to move from any given branch to any other branch generally required the traversal of several vertical moves, as there were often no direct links between branches. In order to reach a node on a new branch, then, it was often necessary to first move up to the Gilded Age overview card, and then down to the topic of choice. When it was possible to move to another branch via a lateral move, however, only a single link traversal was often required. All systems, then, not only permitted, but required, much wider use of vertical links.

4.1. Depth of Understanding

One prediction made by this study was that there would be differences between system groups in terms of the depth of learning acquired by participants; that is, how much learners were able to move beyond the factual information presented in the systems to arrive at a more meaningful understanding of the subject matter. The essay results were consistent in showing superior performance of the unstructured participants over the highly structured and linear participants.

How can this pattern of results be explained? The major difference between the highly structured and the unstructured conditions is that the nature of the relations between links was made more explicit in the highly structured case. As a result, it should have been simpler for highly structured participants to stay oriented in the system without putting a tremendous amount of thought into the relations between nodes. For example, they were able to move up branches by pressing arrow buttons without reading the button names. Those in the unstructured condition, however, ran the risk of becoming disoriented (a common problem in hypertext systems) unless they put more thought into their navigation. If they did not concentrate on the link names and where they were going, they ran the risk of getting seriously lost in the system. In short, the less structured system required a deeper level of processing of the information implicitly provided by the links in order to stay oriented in the system and make sense of the material. The more organized nature of the hierarchical link structure mitigated the necessity of deeply processing the information embedded in the links. Participants in the highly structured condition were able to move through the information less thoughtfully than those in the unstructured condition.

The navigation trail results support this explanation. As discussed in the results section for the navigation data, the highly structured participants tended to use many more lateral links than those in the unstructured group. As illustrated in Figure 1, superordinate nodes were located on a separate part of the document from lateral links. As a result, the lateral links were more handy when a user got to the end of a document. It would appear, then, that the highly structured group used ease of access as a major criterion for link choice. In other words, they were highly influenced by the structure of their system. Because all links appeared equivalent to the unstructured group and this group also chose a greater variety of links to follow, it is likely that members of this group put more thought into their behavior.

Putting the learner in a position to actively make the content of a text understandable and cohesive has been shown to improve learning outcome in many studies of text-based learning (Craik & Tulving, 1975;

O'Brien & Myers, 1985). In discussing a review of the literature on retention strategies for text, Mannes and Kintsch (1987) noted that refraining from "providing readers with a suitable schema and thereby forcing them to create their own ... might make learning from texts more efficient" (p. 93). In one such study, Dean and Kulhavy (1981) presented learners with a text on an imaginary African tribe. Half the participants were asked to draw a relational map of the territory and events depicted in the article. Posttests required integration rather than simple recall; they were designed to test the situation model. Participants in the map condition, who were put in a position to actively explore the meaning and relations between events in the passage, outperformed their counterparts in the no-map condition on each posttest.

This result is similar to one reported by McNamara et al. (1996), discussed earlier, that less coherent texts actually resulted in better comprehension for high-knowledge participants. The authors of that study contended that the act of filling in gaps in the text resulted in the construction of a robust situation model because individuals were forced to actively construct meaning on their own if they were to understand the text. It is reasonable to conclude that those participants in this study who were also put in a position to be thoughtful about the coherence of the system as a whole (by coming to an understanding of the link relations) underwent a similar process. The analyses here suggest, then, that there are strong similarities between traditional text comprehension and hypertext comprehension. As discussed in the previous section, results indicate that text bases are formed for the information contained on documents and for the existence of conceptual links. However, integrating the information represented by those links would seem to require a deeper level of processing; the construction of a situation model.

There is one caveat to this conclusion, however. McNamara et al. (1996) found that the benefit of a low-coherence text was only gained by high-domain-knowledge learners; those with low domain knowledge benefited more from the more highly structured text. All participants in this study were low domain learners, as exemplified by the pretests they all took (and failed in order to qualify for the study), but performed better overall with the lower structured system. The lack of parallel findings suggests one of two possibilities. One is that a high-coherence text, as defined by McNamara et al., is not equivalent to a highly structured hypertext system, as defined here. The other possibility is that there are fundamental differences between text and hypertext learning that this investigation was unable to discern.

Clarifying this question is a matter for future investigation. However, the explanation favored by the navigation trail data is that the way in which learners approached the system may be important. Specifically, the way in which participants navigated through the material hints that the global structure of a hypertext system may be less relevant to the learning outcome than the way in which students use the system. Certainly, user-friendly systems offer many advantages to users (Edwards & Hardman, 1989; Hammond & Allinson, 1989; Laurel, Oren, & Don, 1990; Marshall & and Irish, 1989; Nielsen, 1989, 1990). However, this study suggests that if a highly structured system is counterproductive, it may be because it offers an easier way to get around the system without requiring much thought.

4.2. Factual Knowledge

Another prediction made for this investigation was that the structure of a hypertext system would have no bearing on the amount of factual information participants learn. Because the factual information presented by each of the study's systems was invariant, there should be no difference in learning outcome despite differing interfaces. In fact, there were no differences between groups on the short-answer posttest.

In addition, the concept maps of both experimental groups included the same amount of factual information. The linear control group did show a difference on this measure but these learners were never exposed to any links at all. If this result may be understood in terms of the creation of text bases, both linked systems, regardless of their structure, were equivalent in their ability to impart knowledge about the existence of associations between ideas. This result is an indication, however, that learners working with hypertext systems readily acquire the factual information to which they are overtly exposed—even the information provided within the link buttons.

4.3. Learning From Hypertext

The final goal of this inquiry was to learn more about the general effectiveness of hypertext systems. Results already presented here address this issue. Specifically, although there were differences between experimental groups on posttests of integration and understanding, neither of these groups outperformed the linear group. This result puts the effectiveness of hypertext over that of more traditional text into question. However, the fact that the unstructured system proved to be superior to the highly structured system indicates that the design of a hypertext system does have an effect on learning. Although neither of the hypertext systems in this study have shown an advantage over learning from traditional text, it is highly possible that the reason may stem from some shortcoming of the systems developed for these purposes. Indeed, other studies have shown augmented educational outcomes due to hypertext (Chen & Rada, 1996).

Another important point to emerge from the navigation logs is that participants working with a linked hypertext system miss documents. Participants in the linear group saw significantly more of the information than the other two groups because they were not able to skip documents in the linear sequence. This result highlights a potential problem with learning from hypertext. That is, it is possible for users to miss whole portions of the information.

The concept map results provided more positive information about learning from hypertext. As mentioned earlier, the concept map analyses showed that the linear group included fewer links on their maps that were programmed into the two experimental systems. Obviously, it makes sense that the linear group would include less information about links to which they were not exposed. This result is important, though, because it highlights the fact that both experimental groups were affected by the presence of the system links. Both linked system groups remembered the associations between linked ideas, incorporated those associations into their understanding of the material, and were able to illustrate that knowledge in their concept maps. If that were not the case, the links they included in their maps would not have differed from those of the control group. This result points to a potential benefit of learning from hypertext rather than traditional text. That is, information about relations between ideas is readily acquired by learners when they are programmed into system links. This is a benefit that traditional text simply cannot offer. It is possible, of course, to list the names of related topics alongside text within traditional text-based documents. However, results from an unrelated study in my laboratory indicate that participants do not benefit from their presence (Shapiro, 1997). It appears that mere exposure to link button names may be different from actually working with the links.

5. CONCLUSIONS

The first aim of this study was to explore the hypothesis that theories of learning from text may be transferred to learning from hypertext. Figure 6 displays the major findings that generally support this idea. Current models of learning from text assert that shallow learning takes place when information situated within the text is incorporated into memory. Deeper learning occurs when that information is integrated with prior knowledge. To obtain this deeper knowledge, it is necessary to engage in more involved processing than that required for acquisition of mere facts from the text itself. If hypertext learning is similar to text-based learning, it should be the case that deeper processing produces deeper learning. With this in

***Figure 6.* Summary of the study's main results.**

Analysis	System HS	US	L	Interpretation
Integration				
Essay performance	Low	High	Med	The US group was better able to integrate the information and present it in their essays.
Concept maps	Med	Med	Med	No significant effect of system structure on this measure.
Factual knowledge				
Concept map nodes	High	High	Low	Both hypertext groups included significantly more nodes in their maps than the L group. They had more information available to them to include in the maps.
Link familiarity	High	High	Low	Both hypertext groups included significantly more links classified as "familiar" in their maps than the L group. The links in these systems were useful for imparting information about topic relations.
Short-answer	Med	Med	Med	All systems were equally efficient at imparting factual information from the documents.
Essay density	Med	Med	Med	All participants were equivalent in their ability to recall and include factual information in their essays.
Navigation				
Amount read	Med	Med	High	The linked systems offer the possibility of "missing" information because the user must choose which documents to visit.
Time spent	Med	Med	Med	No significant effect of system structure on this measure.
Navigation pattern	High % of lateral moves	Low % of lateral moves	N/A	The US group used a more thoughtful approach to navigation than the HS group.

Note. HS = highly structured; US = unstructured; L = linear.

mind, the three systems created for this study presented learners with varying opportunities to process the information. Specifically, the unstructured system provided little information about topic relations and so provided the greatest opportunity for increased processing of the information. The highly structured system provided more information about topic relations.

The measures of integration used in this study supplied mixed results. Although the concept map results showed no difference between system conditions, there was a significant effect on learners' essays. Specifically, those in the unstructured condition produced the highest quality essays. This result is taken to mean that they did engage in a deeper level of processing. The navigation results that provided information about this group's strategy for exploration support this conclusion. The lack of significant results on the concept map link density measure could mean that learners' conceptual structures were not strongly affected by the different systems. In light of the long history of research that has tied conceptual structure to learning outcome, however, I am inclined to conclude that link density is an ineffective way of measuring the complexity of participants' conceptual structures. Other methods such as cued topic association and card sorting have been used more profitably by others (McNamara et al., 1996; Shapiro, 1997).

The construction integration model of text-based learning also predicted that there should be no difference between hypertext groups with reference to the amount of factual knowledge each is able to acquire. This is so because the factual knowledge presented on each document is invariant across systems. As shown in Figure 6, there were no differences between hypertext groups on any of these measures. The only significant differences observed between groups in factual knowledge acquisition was between the linear group and the hypertext groups. These differences, however, all stemmed from measures of knowledge about the links, to which the linear group was never exposed. Those differences, then, are not particularly relevant to the application of theories of learning from text to learning from hypertext. They do, however, shed light on the second topic of this inquiry, the educational value of hypertext.

Any explicit, factual information that was presented equally to the three groups in the study was learned equally by participants in each condition. In this arena, learning from hypertext is similar to learning from text. In addition, all hypertext groups performed comparably on all four of these measures. As noted earlier, the only significant differences between groups surfaced between the linear group and the hypertext groups on measures related to the surface information contained on the links. The linear group was used as a control in these cases because they obviously had no exposure to the links as they worked. These results indicate that surface information presented by the existence of hypertext links (i.e., the fact that there are relations between linked documents) is indeed acquired by learners. As such, the results speak to the potential effectiveness of hypertext over traditional text.

The significant effect of system condition on essay performance also indicates that hypertext can have educational value over traditional text, as the unstructured group significantly outperformed the highly structured group. This group also performed slightly better than the linear group, although the difference was nonsignificant, as was the difference between the highly structured and linear groups. This pattern of results indicates that the structure of a hypertext system is crucial to the system's educational value. Simply put, suboptimal hypertext design will diminish the effectiveness of hypertext-based instruction, potentially to a point below that of traditional text. The structure of a system may in fact alter the learning outcome by affecting the way in which users approach the materials in a hypertext system. After all, learners working with the unstructured system navigated differently (more thoughtfully) from those in the highly structured condition and performed better on all measures of the essay test and the highly structured group actually performed slightly worse than the linear group (although the difference was nonsignificant). Because neither hypertext group performed better than the linear group, this pattern of results fails to provide definitive evidence that hypertext is of greater benefit for this type of learning than traditional text. This conclusion is consistent with Chen and Rada's (1996) meta-analysis of hypertext-based learning, in which they reported mixed results regarding the benefit of the technology over text. In their paper, Chen and Rada concluded that system design may be one important factor in determining the value of hypertext. This conclusion is supported by the differences in learning outcome between the hypertext groups in this study.

What elements of system design are required to make the technology more educationally effective? Learners in this study acquired the surface information presented by the links regardless of system structure, their navigation behavior, and so forth. It is the acquisition or creation of the deeper meaning behind this surface information (i.e., the relations between linked documents) that is less certain with this medium. The results reported here would seem to indicate that system design is important because it may be tied to patterns of use on the part of the learner. If this is an important consideration, the design of the system should be geared toward guiding profitable user behavior.

How should learners be approaching the use of a hypertext system? Obviously, this will depend on the goal of the learner. However, if the goal is to acquire an integrated, deep understanding of the material presented in the system, the results presented here indicate that more active processing of the information is necessary. Specifically, links must be used as more than vehicles for navigation or mere pointers to conceptual ties. A degree of thoughtfulness must be given to these relations, as well. It is certainly not suggested here that system designers create unwieldy and confusing products for educational purposes. Rather, the challenge facing system designers and educators interested in the benefit of hypertext-aided instruction will be to find ways to make navigation simple and require students to be thoughtful about the relations implied by document links.

The tension lies in finding the right balance between challenge and system cohesion. It is highly unlikely that this balance would be the same for every student, which makes the issue more complicated.

Incorporating components of intelligent tutoring systems could be one solution. Programs such as those created by Anderson and colleagues at Carnegie Mellon University (Anderson, Boyle, Farrell, & Reiser, 1987) have had great success in assessing students' expertise level and using that information to customize the information offered. This is done by monitoring students' successes and failures and creating internal "lists" of misconceptions and correct rules of which the student has displayed knowledge. This information is used by the system to choose level-appropriate problem sets or hints for the student. As Hammond (1990) noted, however, this type of very directed guidance has its limits in that it removes control from the user. User control is, of course, one benefit of hypertext. However, researchers have advocated the use of some form of hybrid system that incorporates a hypertext web with intelligent components such as adaptive links (Hammond, 1990; Tyerman & Verbyla, 1991).

For example, Tyerman and Verbyla (1991) described a system they called the hypermedia tutor. The system contains a large network of nodes and links. However, a user-friendly interface allows the instructor to block out whole segments of the system, allowing the user to move freely within only a portion of the corpus. Factors such as a student's performance on an exam may be programmed in to determine which segment of the system the individual is guided toward. Such a solution seems promising because it allows learners to retain control in a way that encourages them to actively explore the information space, a strong point of hypertext. On the other hand, the intelligent component of such a system takes control away from learners only in determining their appropriate level of difficulty. This is important because the kind of metacognitive ability necessary to monitor one's own understanding is notoriously lacking in most students and is a difficult skill to acquire, especially for children. By taking control of this difficult aspect of learning from the user while preserving the user's autonomy in exploring the appropriate information space, the process of active learning may take place unimpeded. However, research discussed in the beginning of this inquiry also points to the dangers of giving unhindered control of even this sort to learners. It seems that users often need guidance in approaching hypertext material in an active and productive manner.

Another approach to this problem, then, may be to step outside the domain of system design and concentrate on the skills of the users. Landow (1992) stated that "[c]ritical thinking relies upon relating many things to one another" (p. 126). He went on to say that hypertext provides a means of "accustoming students to making connections among materials they encounter" (p. 126). Indeed, it is almost certain that, like most study skills, critical thinking in this context must be taught to students, especially younger learners who may not be developmentally able engage in the kind of problem solving necessary to understand the abstract relations between linked documents. Research currently underway in my laboratory is directed at developing an instructional system, based on the cognitive apprenticeship model (Collins, Brown, & Newman, 1989), for teaching students strategies for approaching hypertext materials in a thoughtful way. The focus of the program is to teach the students how to figure out the relations between linked documents as they work. This will be done by first modeling the behavior for them and progressively giving control to them as they attain more practice.[7]

Whatever the outcome of this approach or any other, it is important to better understand the process of learning from hypertext if the medium is to become a robust and dependable teaching tool. It is only through this avenue that effective system designs and instructional strategies will be developed. At present, there is no widely accepted framework for system design and no proven instructional strategy. This study was conducted in the hope of contributing to our understanding of how we learn from hypertext. However, the use of the technology in educational settings is still vastly unexplored. As discussed earlier, many issues related to essential elements of system design and optimal user strategies are still unknown and ripe for exploration.

NOTES

Acknowledgments. I gratefully acknowledge the help of Kathryn Spoehr, Sharon Carver, Lynne O'Brien, and Joseph Campione for their help on the various incarnations of this manuscript. Many thanks also to two anonymous reviewers for their advice, which greatly improved the analyses and content of this work. Finally, I would like to thank Eileen and Walter Kintsch for their generous time in offering perspective on the analyses.

Support. This research was funded by a James S. McDonnell Foundation grant to Kathryn Spoehr and preparation of the manuscript was generously supported by a James S. McDonnell Foundation fellowship to the author.

Author's Present Address. Amy M. Shapiro, Psychology Department, University of Massachusetts Dartmouth, North Dartmouth, MA 02747-2300. E-mail: ashapiro@umassd.edu.

HCI Editorial Record. First manuscript received September 26, 1995. Revision received March 17, 1997. Accepted by Elliot Soloway. Final manuscript received May 5, 1997. — *Editor*

7. This 3-year research project has been funded by the James S. McDonnell Foundation's Cognitive Studies for Educational Practice program.

REFERENCES

Anderson, J. R., Boyle, C. F., Farrell, R., & Reiser, B. J. (1987). Cognitive principles in the design of computer tutors. In P. Morris (Ed.), *Modeling cognition* (pp. 93–113). New York: Wiley.

Beeman, W. O., Anderson, K. T., Bader, G., Larkin, J., McClard, A. P., McQuillan, P., & Shields, M. (1987). Hypertext and pluralism: From lineal to nonlineal thinking. *Hypertext '87,* 69–85. New York: ACM.

Bower, G. H., Clark, M. C., Lesgold, A. M., & Winzenz, D. (1969). Hierarchical retrieval schemes in recall of categorized word lists. *Journal of Verbal Learning and Verbal Behavior, 8,* 323–343.

Chen, C., & Rada, R. (1996). Interacting with hypertext: A meta-analysis of experimental studies. *Human–Computer Interaction, 11,* 125–156.

Collins, A., Brown, J. S., & Newman, S. E. (1989). Cognitive apprenticeship: Teaching the crafts of reading, writing, and mathematics. In L. B. Resnick (Ed.), *Knowing, learning, and instruction: Essays in honor of Robert Glaser* (pp. 453–494). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Craik, F. I. M., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior, 11,* 671–684.

Craik, F. I. M., & Tulving, E. (1975). Depth of processing and the retention of words in episodic memory. *Journal of Experimental Psychology: General, 104,* 268–294.

Dean, R. S., & Kulhavy, R. W. (1981). Influence of spatial organization on prose learning. *Journal of Educational Psychology, 73,* 57–64.

Edwards, D. M., & Hardman, L. (1989). "Lost in hyperspace": Cognitive mapping and navigation in a hypertext environment. In R. McAleese (Ed.), *Hypertext: Theory into practice.* Oxford, England: Intellect.

Eylon, B., & Reif, F. (1984). Effects of knowledge organization on task performance. *Cognition and Instruction, 1,* 5–44.

Hammond, N. (1990). Teaching with hypertext: Problems and prospects. In H. Brown (Ed.), *Hypermedia, hypertext, and object-oriented databases* (pp. xxx–xxx). London: Chapman & Hall.

Hammond, N., & Allinson, L. (1989). Extending hypertext for learning: An investigation of access and guidance tools. In A. Sutcliffe & L. Macauley (Eds.), *People & computers* (pp. 293–304). Cambridge, England: Cambridge University Press.

Hulse, S., Deese, J., & Egeth, H. (1975). *The psychology of learning* (4th ed.). New York: McGraw-Hill.

Kintsch, W. (1988). The use of knowledge in discourse processing: A construction integration model. *Psychological Review, 95,* 163–182.

Kintsch, W. (Ed.). (1974). *The representation of meaning in memory.* Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Landow, G. (1992). *Hypertext: The convergence of contemporary critical theory and technology.* Baltimore: Johns Hopkins University Press.

Laurel, B., Oren, T., & Don, A. (1990). Issues in multimedia interface design: Media integration and interface agents. *CHI'90 Proceedings,* 133–139. Seattle, WA: ACM.

Leggett, J. L., Chase, J. L., & Kacmar, C. J. (1990). Hypertext for learning. In D. H. Jonassen & H. Mandl (Eds.), *Designing hypertext/hypermedia for learning*. Heidelberg, Germany: Springer-Verlag.

Leonard, J., & Whitten, W. (1983). Information stored when expecting recall or recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 9,* 440–455.

Light, L., & Carter-Sobell, L. (1970). Effects of changed semantic context on recognition memory. *Journal of Verbal Learning and Verbal Behavior, 9,* 1–11.

Mannes, B., & Kintsch, W. (1987). Knowledge organization and text organization. *Cognition and Instruction, 4,* 91–115.

Marchionini, G. (1989). Information-seeking strategies of novices using a full-text electronic encyclopedia. *Journal of the American Society for Information Science, 40,* 54–66.

Marshall, C. C., & Irish, P. M. (1989). Guided tours and on-line presentations: How authors make existing hypertext intelligible for readers. *Hypertext '89 Proceedings,* 15–26. Pittsburgh, PA: ACM.

McGrath, D. (1992). Hypertext, CAI, paper, or program control: Do learners benefit from choices? *Journal of Research on Computing in Education, 24,* 513–531.

McNamara, D. S., Kintsch, E., Songer, N. B., & Kintsch, W. (1996). Are good texts always better? Interactions of text coherence, background knowledge, and levels of understanding in learning from text. *Cognition & Instruction, 14,* 1–43.

Meyer, B. J. F. (1975). *The organization of prose and its effects on memory*. Amsterdam: North-Holland.

Meyrowitz, N. (1986). Intermedia: The architecture and construction of an object-oriented hypertext/hypermedia system and applications framework. *OOPSLA '86 Proceedings,* 186–201. Portland, OR: ACM.

Murphy, G. L., & Shapiro, A. M. (1994). Forgetting of verbatim memory in discourse. *Memory & Cognition, 22,* 85–94.

Nielsen, J. (1989). The matters that really matter for hypertext usability. *Hypertext '89 Proceedings,* 239–248. Pittsburgh, PA: ACM.

Nielsen, J. (1990). The art of navigating through Hypertext. *Communications of the ACM, 33,* 296–310.

O'Brien, E. J., & Myers, J. L. (1985). When comprehension difficulty improves memory for text. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 11,* 12–21.

Shapiro, A. M. (1997). *The relevance of hierarchies to learning from hypertext*. Unpublished manuscript.

Spoehr, K. T., & Shapiro, A. M. (1991, April). *Learning through hypertext: Making sense of a multiply-linked database*. Paper presented at the annual meeting of the American Educational Research Association, Chicago.

Tyerman, S. P., & Verbyla, J. L. (1991). Intelligent computer aided instruction using hypertext. *Proceedings of the 8th Annual Conference of the Australian Society for Computers in Leaning in Tertiary Education,* 639–647. Launceston, Tasmania: ASCILITE.

van Dijk, T. (1980). *Macrostructures: An interdisciplinary study of global structures in discourse, interaction, and cognition*. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.