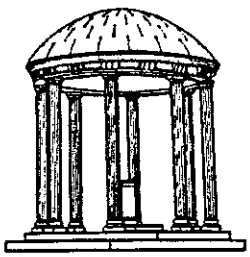
Using Computer-Generated Protocols to Study Writers' Planning Strategies

TR90-033

September, 1990

Marcy Lansman
John B. Smith
Irene Weber



The University of North Carolina at Chapel Hill
Department of Computer Science
CB#3175, Sitterson Hall
Chapel Hill, NC 27599-3175

A TextLab Report

Portions of this research were supported by the National Science Foundation Grant #IRI-8519517 and the Army Research Institute Contract #MCA 903-86-C-0345.

Abstract

Research has shown that extensive planning is typical of skilled adult writers (Bridwell-Bowles, Johnson, & Brehe, 1987; Flower & Hayes, 1981b). Yet word processors tend to discourage high level planning (Haas, 1989). In this research, we used the Writing Environment, a computer-based authoring tool that encourages planning, to study the strategies of adult writers. Nine graduate students and eight technical writers used the Writing Environment to write technical reports, while the computer recorded their activities. Analysis of computer-generated protocols revealed that subjects spent a large proportion of their time on the organizational structures for their reports and that these structures were quite elaborate. Subjects varied widely in the extent to which they completed their plans before they wrote. Surprisingly, the overall quality of the reports, as rated by two judges, was negatively related to time spent planning. Some writers apparently spent too much time on complex organizational structures and too little time composing text.

Using Computer-Generated Protocols to Study Writers' Planning Strategies

People who study writing agree that planning is a good thing: that mature writers plan more than immature writers; that writers who use outlines write better compositions than writers who compose without outlines; and that expert writers plan more than novice writers. Yet research on word-processors has shown that writers who use computers plan less than writers who use other methods. Standard word processing programs encourage sentence-level composing and editing while discouraging more global planning and revising. The purpose of the research reported here was to introduce an authoring tool that facilitates planning and to study the planning strategies of adult writers as they composed a technical report using that tool.

Studies of Planning

How researchers study planning depends on how they define it. Planning entails mental processes such as generating ideas and setting goals. It also involves the physical process of recording those ideas and goals. Some researchers have focussed on written plans, while others have been more interested in the cognitive processes involved in planning.

Research on written plans. The ability to generate written plans seems to develop relatively late in childhood. Bereiter and Scardamalia (1987) noted a dramatic difference between the written plans of children 10 to 14 and those of college students. The children's plans were almost identical to their final texts. In spite of explicit training, children failed to engage in "conceptual planning" as distinct from "content generation," to use Bereiter and Scardamalia's terminology.

Left to their own devices, even high school and college students rarely produce written plans. Detailed studies of the writing habits of high school and college students have shown that although these students have been taught outlining and are often required to turn in outlines at school, they do not find written plans useful and do not choose to create them unless required to do so (Emig, 1971; Mischel, 1974; Pianko, 1979; Stallard, 1974). But observations of experienced adult writers have shown that they often spend a large proportion of their time creating elaborate written plans, sometimes in the form of outlines and sometimes in the form of notes, lists and diagrams (Berkenkotter, 1983; Bridwell-Bowles, Johnson, & Brehe, 1987; Selzer, 1983).

Kellogg has used both observational and experimental studies to demonstrate that planning helps adult writers. In a survey of science and engineering faculty, he found that those professors who made greater use of written plans were more productive (Kellogg, 1986). In experimental studies, Kellogg also asked college students to write short business letters or essays with or without outlining (Kellogg, 1987, in press). The subjects

who outlined first produced significantly better texts than those who began writing immediately.

Thus several lines of evidence support the conclusion that written plans are typical of mature, successful writers and lead to better written products.

Research on mental plans. Recent concern with the writing process, as opposed to the written product, has caused many researchers to focus on mental rather than physical plans (Faigley, Cherry, Jolliffe, & Skinner, 1985; Hagge, 1987; Hairston, 1982). To study mental planning processes, they have used think-aloud protocols, generated when writers verbalize their thoughts as they compose (Swarts, Flower & Hayes, 1984).

On the basis of think-aloud protocols, Flower and Hayes (1981b) have argued that planning is central to adult writing. They have defined a broad variety of cognitive activities as planning: goal-setting, audience analysis, idea generation, organization, analysis of the rhetorical problem, and more. Whereas earlier models of writing conceived of planning as a separate stage that precedes writing and revising, Flower and Hayes argued that planning is distributed throughout the writing process. Furthermore, they claimed that written plans reveal only a small proportion of the planning that goes on inside writers' heads (Flower & Hayes, 1981b).

An especially important aspect of planning, according to Flower and Hayes, is elaboration of the rhetorical problem presented by the writing task. Drawing an analogy between writing and problem solving, they pointed out that writers differ in their internal representation of the writing assignment. On the basis of think-aloud protocols, Flower and Hayes (1980) have claimed that expert writers spend more time defining the task than novice writers. Defining the task includes such activities as describing the audience, deciding what effect the text should have on that audience, devising a strategy for achieving that effect, etc. (Flower & Hayes, 1980; see also Stotsky, 1990).

Supporting the claim that good writers plan more than poor ones, Carey, Flower, Hayes, Schriver, and Haas (1987, cited in Hayes, 1989), found a significant positive relationship between both the quantity and the quality of planning clauses in think-aloud protocols of adult writers and the quality of the final texts they produced.

Thus research on think-aloud protocols has corroborated research on written plans: expert writers plan more than novices and writers who plan more tend to produce higher quality text.

The effects of word processors on writing. In most of the research on planning, writers have used paper and pencil to compose. During the past

decade, however, there has been a revolution in writing technology. More and more writers, both experts and novices, are using word processors. They not only enjoy writing more when they use the computer, but are convinced that the machine improves their writing. Participants in formal studies have expressed similar enthusiasm for the word processor (Gould, 1981; Haas & Hayes, 1986), but these studies have failed to demonstrate that word processors improve the quality of users' written texts.

In a wide range of studies, word processors have been compared to more conventional writing tools. For example, Daiute (1986) found that junior high school students wrote better first drafts of letters using paper and pencil than word processor, but that the quality of final drafts was equal. Hawisher (1987) found that freshman composition students produced final drafts of equal quality whether they used word processors or paper and pencil to make revisions. Haas (1989) asked college professors to write and revise letters using either advanced work stations, conventional word processors or paper and pencil. She found that subjects produced higher quality texts using work station and paper and pencil than using conventional word processors. Gould (1981) found that professionals who were highly practiced at using a text editor took 50% longer to write a series of letters using the text editor than writing longhand but produced letters of equal quality. Kellogg and Mueller (1989) found that student essays written on a word processor were rated significantly poorer in style and no different in quality of content than essays written in longhand.

Although few studies, if any, have shown that word processors improve text quality, many have shown that they affect writers' strategies. A study by Bridwell-Bowles, Johnson and Brehe (1987) described these effects in detail. They asked eight writing instructors who had no experience with word processors to complete four assignments, one using their usual methods and three using word processors. Using their accustomed methods, all of these writers produced elaborate written plans. However, they differed in the extent to which they planned before they wrote. At one extreme, "Mozartians" completed their planning before they began to write. At the other extreme, "Beethovians" discovered what they wanted to say as they wrote. All the writers found the word processor more useful for writing and revising than for planning, but the Beethovians were particularly frustrated in their attempts to use the computer to plan.

Other researchers have noted that the word processor encourages sentence-level revisions but discourages planning and large scale revision. Lutz (1987), for example, asked professional writers and graduate students to revise their own and others' texts either by hand or using the word processor. She found that they worked longer and made more changes using the word processor, but that the changes made on the word processor tended to be at a more local level. Haas (1989) asked college professors to revise their own letters using either work station, conventional word processor, or paper and pencil. She found that when they used paper and

pencil they spent more time planning and concentrated their planning before they began the actual revision. Gould (1981) found that professional researchers made many more revisions to their letters when composing on the text editor than in longhand. Collier (1983) looked at videotapes of four freshman composition students revising their papers using either word processor or paper and pencil. He observed that subjects engaged in more sentence-level revision using the word processor. In his words: "Analysis of the videotaped session revealed that revision was more complexly layered when it was performed on a word processor: revisions were themselves often revised; revisions were enacted and then deleted (sometimes several times)" (p. 151).

These studies lead to the disturbing conclusion that while planning is a critical component of the writing process, traditional word processors discourage writers from planning. This finding may explain why word processors have not been shown to have a positive effect on text quality.

A New Authoring Tool

In response to the need for a computer tool that supports planning, we have developed an experimental program called the Writing Environment (Smith, Weiss, Ferguson, Bolter, Lansman, & Beard, 1987). The program is built on the assumption that writers engage in several modes of thinking, each involving its own goals, processes, and products (Smith & Lansman, 1989). For example, idea generation is one mode of thinking employed by writers. In idea generation the goal is to generate information to be included in a text; the cognitive processes are search of long-term memory and written materials; and the product is a set of loosely related ideas. Other cognitive modes include organization, sentence generation, global editing, and local editing.

The conventional word processor directly supports only sentence generation and local editing. The Writing Environment, on the other hand, supports several other modes of thinking by including four separate system modes that appear in four windows on the computer screen. Network Mode is designed for idea generation and exploration, Tree Mode for organization and global editing, Edit Mode for sentence generation and local editing, and Text Mode for coherence editing. By providing separate system modes for the various cognitive modes of writing, the Writing Environment allows writers to focus attention on each of these modes exclusively and to move easily among them.

Of particular concern here are the system modes that support planning: Network and Tree Modes. A number of commercial software packages (such as THINKTANK, MORE and the outlining functions of Microsoft WORD) have been developed to help writers create outlines. The Network and Tree Modes of the Writing Environment differ from these packages in two important respects. First, they are graphics based. They show developing structures as diagrams rather than conventional outlines,

and they permit the writer to manipulate those structures directly, using the mouse. These diagrams use the two dimensions of the screen more flexibly than outlining programs. Second, Network and Tree Modes allow for two phases of planning: a free-form, brainstorming phase, supported by Network Mode, and a more constrained organizational phase, supported by Tree Mode. (The details of these modes are described more fully in the Methods Section below.) Conventional outline programs support only the more constrained organizational phase.

For research purposes, a particularly important difference between the Writing Environment and commercial word processing and outlining programs is that the Writing Environment records the writer's interactions with the computer. The transcripts produced by the Writing Environment thus constitute an alternative to think-aloud protocols. Unlike think-aloud protocols, however, computer-generated transcripts do not require writers to interrupt their thought processes to tell the experimenter what they are thinking. Furthermore, since they are computer-readable, these transcripts lend themselves to high-speed, objective computer-based methods of analysis.

Purpose of this Study

The purpose of this initial study using the Writing Environment was to study writers' planning strategies. Since the Writing Environment was developed to aid adult professionals in their work-related writing, we used adults at two levels of writing experience as subjects. In order to control content knowledge, we assigned them the task of writing a technical report from source materials on an unfamiliar topic.

The study focussed on several dimensions of planning. We wanted to know whether writers would use this new tool to plan their reports, what proportion of their time they would spend planning, and how that time would be distributed across the writing session. We also wanted to know what kind of strategies writers would use to plan and write their reports. Finally, we wanted to know whether individual differences in distribution of time or in planning strategies would be associated with variations in quality of the final reports.

Method

Subjects

There were two groups of subjects. The first group consisted of 9 experienced technical writers, 8 male, recruited through personal contact from a large computer company in the Research Triangle Park near Raleigh, NC. Their ages ranged from 26 to 69 with a mean age of 46. They had worked as technical writers for 3 to 30 years, with a mean of 10 years. The second group consisted of 9 graduate students, 6 male, in the social sciences and humanities at the University of North Carolina, Chapel Hill, recruited through notices and personal contact. Their ages ranged from 24 to 31 with a mean age of 28. All subjects in both groups had previous

experience using word processors. Graduate students were paid $70 for their participation. Technical writers participated in the experiment either as part of their normal work schedules, in which case they were paid by their employer, or during time off from work, in which case they were paid $70.

Materials

The Writing Environment. Subjects wrote their reports using the Writing Environment, which is implemented on a Sun Workstation with a 19-inch, high-resolution screen. As mentioned above, the Writing Environment has four modes of operation. The user chooses a particular mode by moving a mouse-controlled cursor into one of four windows on the screen.

In Network Mode, the user can generate ideas to be discussed in the document. These ideas appear as labeled rectangular boxes, called "nodes," on the screen. The nodes can be moved around the screen with the mouse and can be linked using directional arrows. In Network Mode, the user may place or move nodes freely around the screen and may link any two nodes regardless of the form of the resulting structure.

In Tree Mode, the user can again generate and link idea nodes, but here the links must form a hierarchical structure, and the shape of the hierarchical structure, a right branching tree, is standardized by the program. Nodes, branches, and entire trees can be moved from Network to Tree Mode. Single nodes can be moved from Tree to Network mode.

In Edit Mode, the user can "open up" a node in either Network or Tree Mode and write text that will be attached to that node. Text is written using a rudimentary text editor in which the cursor is again controlled by the mouse. If a node is moved from one position in the structure to another, the text attached to that node moves with it.

In Text Mode, the user can scroll through and edit text attached to a sequence of nodes. The order of the sequence is dictated by the the hierarchical structure built in Tree Mode. An important difference between Edit and Text Mode is that in Edit Mode the writer can create and edit the text attached to only one node at a time. In Text Mode, the writer can view and edit text attached to several nodes and can move text back and forth between nodes.

In Figure 1, the four modes of the Writing Environment are illustrated in the four windows on the screen. In the figure, the windows are sized as they initially appear on the screen, but in fact, the user can resize any window to fill the screen.

Insert Figure 1 about here

A written tutorial (Jenkins, Lansman, & Smith, 1989) instructs users in how to use the four modes of the system and guides them through a series of exercises. The tutorial does not, however, tell the user how the system should be used to create a document.

Articles. All the subjects used the same five articles as source material for the reports that they wrote during the experiment. The topic of the articles was the manufacture and application of shape memory alloys, a group of metal alloys that resume their original shape when heated. The topic was chosen to be unfamiliar yet understandable to all the subjects. The five articles originally appeared in *Time, Business Week, Chemical Week, Light Metal Age,* and *Mechanical Engineering*.

Procedure

Participants came to the lab for two half-day sessions. During the first session they went through the tutorial individually and familiarized themselves with the Writing Environment. It took them about 1.5 hours to learn to use the system. The experimenter was available to answer questions. After completing the tutorial and experimenting with the system, subjects were given a brief test to make sure that they had a basic understanding of how to use it. During the first session, participants also read the source articles on shape memory alloys and filled out a questionnaire on their writing experience.

During the second session, participants wrote a summary report based on the source articles. They were instructed to write the report from the viewpoint of a technical writer in a small manufacturing firm. According to the instructions, the management of the firm was considering expanding into new markets and the writer's task was to summarize the information on shape memory metals.

Participants were required to use the Writing Environment to write their reports. The experimenter was present throughout to answer questions on the system. Source materials and highlighters were available while subjects were writing, but they were not allowed to use paper and pencil. They were told that they could take as much of the half-day session as they needed to write their reports. They were also encouraged to take breaks.

Evaluation

Two judges evaluated each report. Each of the judges was a writer with a Masters degree in English, and each had taught courses in technical writing. They judged the overall quality of the reports on a 5-point scale. They also evaluated the reports along six specific dimensions: concern for

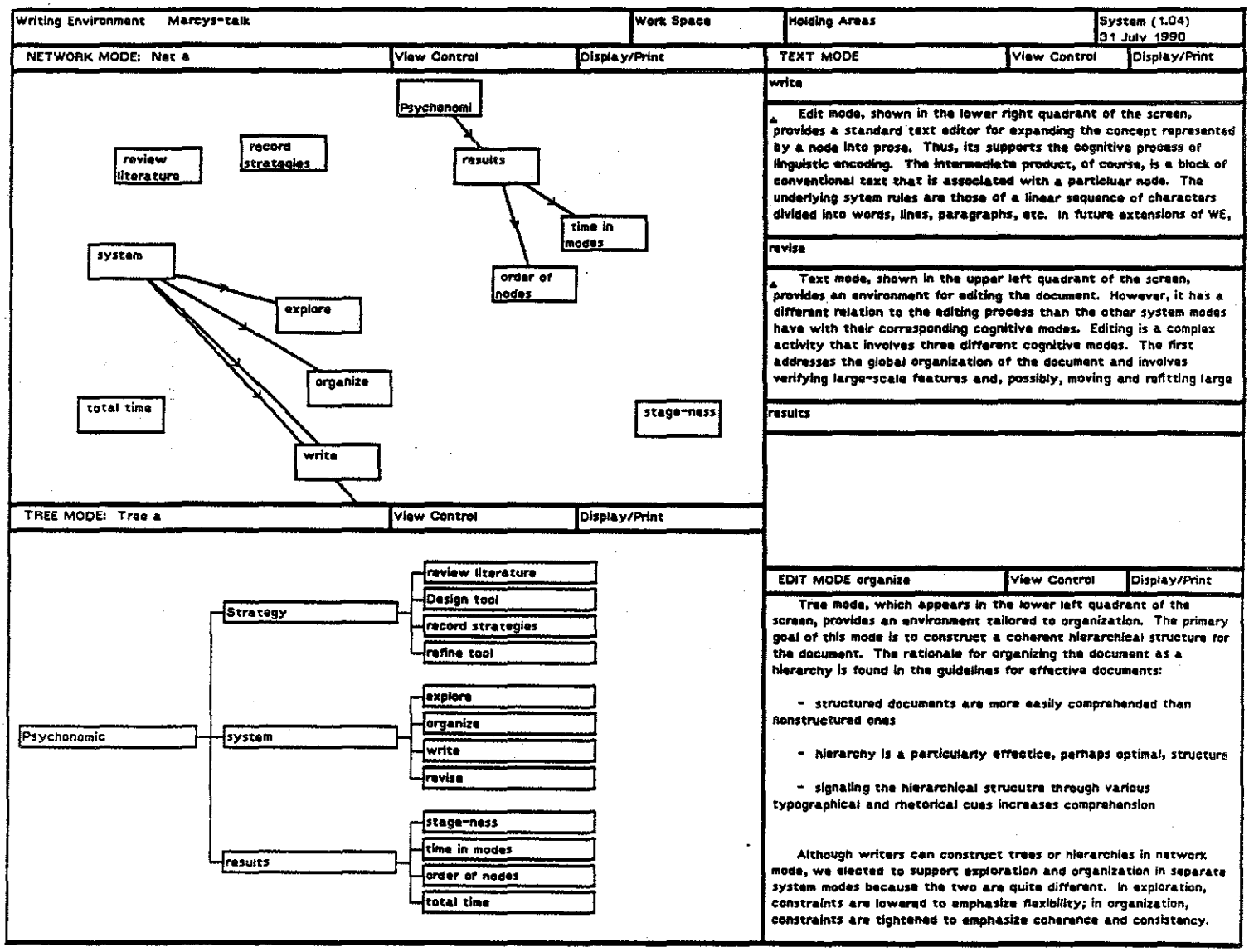


Figure 1. The computer screen might look something like this while a writer was using the Writing Environment to compose a document. Network Mode is shown in the upper left, Tree Mode in the lower left, Edit Mode in the lower right, and Text Mode in the upper right. The user may enlarge any of the modes to fill the screen.

audience and purpose, structure, language, coherence, key points, and accuracy.

Judges participated in four hours of training before they began their evaluations. During the training, they discussed each of the evaluation dimensions and then rated five sample papers along those dimensions. Discrepancies between the judges in their evaluations of the sample papers were discussed and resolved. However, the judges made their evaluations of experimental reports entirely independently.

Results and Discussion

Reports

Each of the subjects but one wrote a report using the Writing Environment. One technical writer left the study at the beginning of the second session explaining that she could not write under lab conditions. The others wrote reports ranging in length from 450 to 1735 words with a mean of 1050. The reports varied in quality from well-organized and polished to quite rough.

Many reports contained an unusually large number of headings. As discussed above, when writers use the Writing Environment, they create text by "opening up" a node in either Network or Tree Mode and composing sentences that will be attached to that node. When the document is printed, the node label appears as a heading which precedes the associated text. Thus each node in the organizational structure becomes a heading in the final document. In these reports, the number of headings ranged from 4 to 46, with a mean of 16.2. Thus there were, on the average, only about 65 words (six lines) of text associated with each heading. For many headings, there was no text at all, indicating that the writer had created a node but failed to write text for it.

The fact that the number of headings was so large relative to the length of the final reports suggests that writers devoted a large amount of effort to planning the organizational structures of their documents relative to the amount of effort spent composing and revising text.

Subjects' Comments

At the end of the second session, subjects were asked to comment in writing on what they liked and disliked about the Writing Environment. On the positive side, subjects were enthusiastic about the planning modes of the system. They felt that the spatial representation helped them develop the structure of their reports. Some thought that Network and Tree Modes were redundant and that one or the other might be eliminated, but they did not agree on which one. On the negative side, subjects complained about the word processing capabilities of Edit and Text Mode, which lacked some of the capabilities of their own word processors.

Evaluations

The two judges were moderately consistent in their ratings of overall quality, r = .67. They were less consistent in their ratings of specific dimensions. (For concern for audience, r = .35; for structure, r = .34; for language, r = .09; for coherence, r = .41; for coverage of key points, r = .46; and for accuracy, r = -.16.) The average of the two judges' overall quality ratings was used as the primary criterion of the quality of the reports.

According to this criterion, the graduate students wrote significantly better reports than the professional writers. On a 1 to 5 point scale, the quality ratings of the graduate students ranged from 1.5 to 4.5 with a mean of 3.39. The quality ratings for the technical writers ranged from 1.0 to 3.0 with a mean of 2.06. This difference was statistically significant, $t = 3.1$, $df = 15$, $p < .01$.

We had expected that the professional writers, with their years of experience writing technical documents, would do better on this task. Clearly, they did not. We do not know whether their inferior performance was due to unfamiliarity with the Writing Environment, unfamiliarity with the topic, or some other factor.

Transcripts

As writers worked on their reports, the computer recorded their activities. The result was a computer-generated transcript, called an Action Transcript, for each subject. The Action Transcript shows writers' activities at a detailed level; e.g., it recorded each time the writer raised a menu, selected an option (e.g., "Create Node"), labeled a node, etc.

The Action Transcript made it possible for us to replay each writing session. During replay, we could see on the screen the same displays that writers had seen as they wrote. Replaying a session showed when and where on the screen writers had created, deleted, and linked nodes in Network and Tree Modes and how they labeled those nodes. It also showed when writers had opened nodes to write text. However, it did not show the text that was written or revised. Thus the replay function provided detailed information on planning for individual subjects but less complete information on writing and revising.

As a first step in the quantitative analysis, a computerized grammar (Smith, Rooks, & Ferguson, 1989) was used to condense the Action Transcript into a less detailed list of activities called the Operation Transcript. For example, in the Operation Transcript the three actions listed above - raising a menu, selecting the Create Node option, and labeling the node - would be condensed into a single operation. The Operation Transcript provided the raw data for further analysis of how participants spent their time as they wrote their reports.

Planning Strategies

Order of creation versus order of appearance. If writers used the Writing Environment for exploration, i.e., to generate ideas freely and then to try out various organizational schemes, then the order in which ideas were generated might be quite dissimilar from the order in which they appeared in the final text. On the other hand, if writers organized their papers in their heads before they began to use the computer, then order of generation might be quite similar to order of appearance in the text.

Our impression, based on replaying the sessions, was that both the order and the superordinate-subordinate relationships among nodes were often roughly worked out in writers' heads before they were recorded on the computer screen. In order to look at this issue more quantitatively, we computed for each subject the correlation between the order in which nodes were created and the order in which they appeared in the final report. The correlations ranged from -.16 to.99 with a mean of .48. The fact that all but two of the correlations were positive and that the mean was fairly high indicates that the order in which subjects generated topic headings was, in most cases, far from random.

There were no differences between graduate students and professional writers on this measure, nor was there a significant correlation between it and quality of the report ($r = -.07$).

Top-down versus bottom-up generation of ideas. One method of creating a hierarchical structure for a document is to write down superordinate topics first, followed by the subordinate topics that go under them. We will call this a "top-down" strategy. It is probably the most common strategy among writers creating an outline in longhand or on a standard outline processor, since writers generally work from top to bottom and superordinate topics always appear above their subordinates in a standard outline. Another method is to generate subordinate topics first and then to group these subordinate topics and create appropriate superordinate topic headings for each group. We will call this a "bottom-up" strategy. Although the bottom-up strategy gives the writer more flexibility in grouping subordinate topics under superordinate topic headings, the strategy is quite clumsy to carry out using word or outline processors. (Some of us have resorted to moving yellow "post-its" around on our desks.) Using the Writing Environment, a writer can easily use either a top-down or bottom-up strategy.

As a group, our writers were much more likely to follow a top-down strategy. As a quantitative measure of "top-down-ness" we used the percent of nodes (omitting the root node) that were generated *after* their superordinate node. This measure ranged from .5 to 1.0 with a mean of .80.

Technical writers adhered more closely than graduate students to a top-down strategy. The mean top-down score for technical writers was .86 and the mean for graduate students was .74. The difference was marginally significant, $t = 1.83$, $df = 15$, $p < .10$.

Furthermore, those subjects who deviated more from a strictly top-down strategy tended to write higher quality reports, as evaluated by our judges. There was a significant negative correlation between top-down score and quality of the report,
$r = -.54$, $p < .05$.

Distribution of Time

Total time spent writing the reports ranged from 2 hours 10 minutes to 4 hours with a mean of 3 hours 11 minutes. Total time was computed as the time that elapsed between writers' first productive system operation (usually, creating a node) and last "save" (i.e., the last time they saved their work to disk). Thus it did not include any time writers spent reviewing the source articles or mentally planning their reports before they began to plan or write using the computer.

Time Per Mode. The Operation Transcript recorded movement of the cursor from one system mode to another. Thus it was possible to compute the time spent and the number of episodes in each of the four system modes. An episode was defined as a series of operations in one system mode uninterrupted by visits to other modes.

Table 1 summarizes the distribution of time among the four modes for all 17 subjects. It shows the mean and range for time spent in each mode, percent of time spent in each mode, and number of episodes in each mode. The mean times spent in Network and Tree Modes, 30 and 22 minutes respectively, indicate that subjects spent considerable time building the organizational structures of their reports. The mean number of episodes in the various modes indicates that subjects moved often among the modes rather than using them in a strictly sequential manner.

Insert Table 1 about here

Edit Mode was intended to be used for working within individual nodes and Text Mode for coherence editing across nodes. The range of values in Table 1 indicates that some writers spent no time in Edit mode and others spent no time in Text mode. These writers preferred to do all their writing and revising in either Edit or Text mode.

Division of time between planning and writing/revising. Assuming that writers did their planning in Network and Tree Modes, we can estimate the amount of time spent planning by adding time in Network Mode and time in Tree Mode. Similarly, we can estimate time spent

Table 1.

Mean and range of time spent in each mode, percent of total time spent in each mode, and number of episodes in each mode.

	Network	Tree	Edit	Text
Mean time (min)	30	22	79	59
Range	6-68	10-45	0-166	0-161
Mean percent of total time	16%	12%	41%	31%
Range	3-37	6-34	0-82	0-82
Mean number of episodes	12	18	14	9
Range	1-34	8-35	0-32	0-17

writing and revising by adding time in Edit Mode and time in Text Mode. A planning episode was defined as a series of episodes in Network or Tree Mode which was uninterrupted by episodes in either Edit or Text mode. Similarly, a writing/revising episode was defined as a series of episodes in Edit and/or Text Mode uninterrupted by episodes in Network or Tree Mode. Table 2 summarizes the distribution of time between planning and writing/revising for all 17 subjects.

Insert Table 2 about here

In computing planning time as the sum of the time spent in Network and Tree Modes, we defined planning time as time spent generating ideas, grouping them and organizing them into a hierarchical structure for the text. Planning time does not include time spent planning the wording of individual sentences. Given this fairly narrow definition, it is noteworthy that writers spent quite a large proportion of their time planning: an average of 53 out of 191 minutes, or 28%.

Other researchers have used quite different methods to estimate planning time. Kellogg and Mueller (1989), for example, estimated percent of time their college student writers spent planning by interrupting them at random intervals and asking them to classify their current activity as planning, translating, reviewing, or other. By this method, they estimated that their subjects spent about 25% of their time planning. Gould (1980, 1981) studied videotapes of professional researchers writing short business letters and classified all pauses when subjects were not actually composing text as planning. By this method, he estimated that his subjects spent two-thirds of their time planning. Selzer (1983), on the basis of interviews, estimated that his single subject, a professional engineer, spent 80% of his writing time "inventing and arranging" the ideas he would include in his written texts. Since the definitions of "planning" used by these researchers and by our project are all so different, it is difficult to draw meaningful comparisons.

Distribution of planning time. According to a traditional "stages" model of composition, writing should take place in three sequential stages: planning, writing and revising. In this model, then, planning occurs (or at least *should* occur) before writing and revising begins. On the basis of think-aloud protocols, Flower and Hayes (1981a) have concluded that planning, writing, and revising processes are called up recursively rather than serially. By studying how our subjects moved back and forth between the planning modes (Network and Tree) and the writing/revising modes (Edit and Text), we can find out whether their strategies were consistent with a strict stage model of composition, or whether a more flexible model, such as that of Flower and Hayes, is appropriate.

Table 2.

Mean and range of time spent planning and writing/revising, percent of total time spent planning and writing/revising, and number of episodes spent planning and writing/revising.

	Planning	Writing/Revising
Mean time (min)	53	138
Range	22-88	63-194
Mean percent of total time	28%	72%
Range	11-54	46-89
Mean number of Episodes	19	19
Range	6-46	6-45

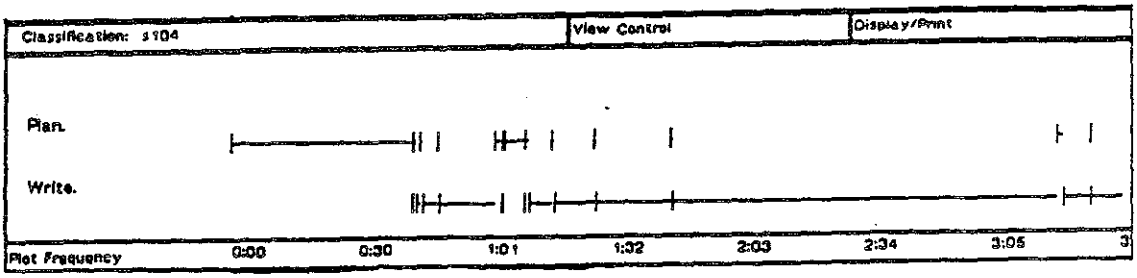
The data from four individual writers are shown graphically in Figure 2, in which planning and writing/revising episodes are represented by the interrupted lines labelled "Plan" and "Write." Figure 2a and 2b show subjects who conformed quite closely to a stage model. They did almost all of their planning before they began to write. Figure 2c and 2d show subjects who used a quite different strategy; they alternated between planning and writing throughout the session. If we consider these two pairs of subjects as defining the two ends of a continuum, the subjects in our study were spread out across the entire range of the continuum with no noticeable clusters. Only a very few could be said to have conformed closely to a stage model. The others all deviated to various degrees from this pattern. Thus there was tremendous variation in the extent to which subjects planned before they wrote. To use the terminology introduced by Bridwell-Bowles et al (1987), our subjects included both "Mozartians" and "Beethovians" and everything in between.

Insert Figure 2 about here

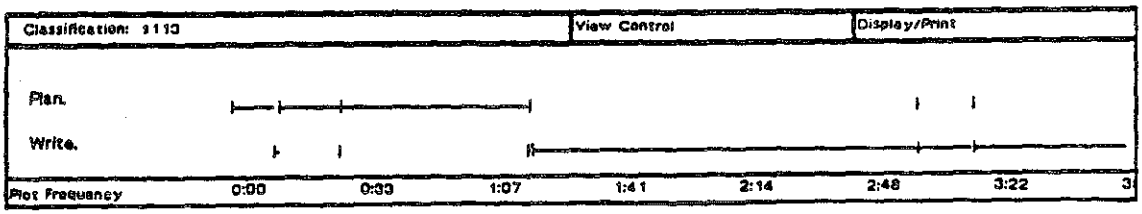
In order to measure more rigorously subjects' conformity to a stage model of writing (i.e., planning first, followed by writing/revising), we developed a Stage Index, which was computed for each subject. This index was designed to assess the extent to which planning time preceded writing/revising time. In order to understand the Stage Index, imagine computing for every minute of writing time the proportion of total planning time that preceded that minute of writing. In order to compute the Stage Index, these proportions are averaged across all the minutes of a writing session. To take the simplest example, if a subject completed all planning before beginning to write, then for each minute of writing the proportion of planning that preceded that minute would be 1.0 and the average, the Stage Index, would be 1.0. The index can vary between close to 0 and 1.0. (It can't be 0 because subjects using the Writing Environment must create at least one node in either Network or Tree Mode before beginning to write.) The Stage Index for the subjects in Figure 2a, 2b, 2c, and 2d was .95, .98, .72, and .67 respectively. For the group as a whole, it ranged from a minimum of .58 to a maximum of .98 with a mean of .78. Although these subjects did not, as a group, conform very well to a stage model of writing, they did concentrate the majority of their planning time before their writing time, as one would expect.

Group differences. There were no statistically significant differences between the graduate students and the professional writers on any of the measures related to distribution of time. There was, however, a tendency for the technical writers to spend a greater percent of their time planning (33%) than the graduate students (24%), $t = 1.74$, $df = 15$, $p = .103$. There were no differences between the two groups in the Stage Index.

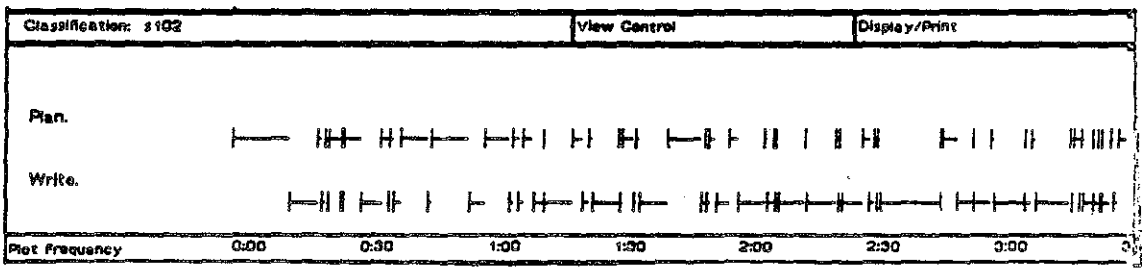
a.



b.



c.



d.

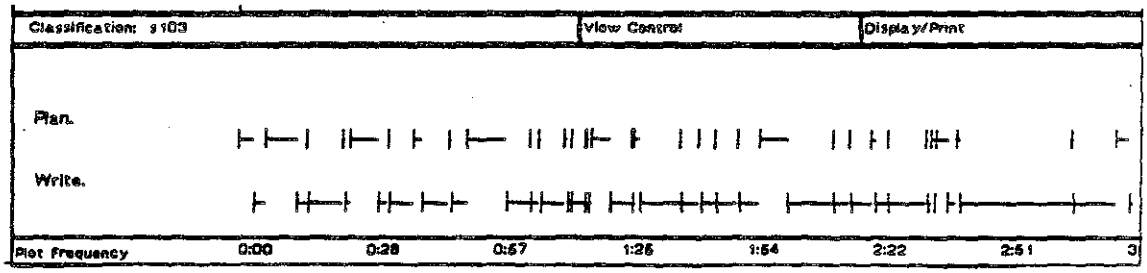


Figure 2. Each panel of this figure shows how an individual subject distributed his or her time between planning (Network and Tree Modes) and writing (Edit and Text Modes). Time since the beginning of the session is shown on the horizontal axis. Each vertical tick represents a planning or a writing episode. The length of the horizontal line attached to the tick represents the duration of the episode. Figures 2a and 2b represent writers who did almost all of their planning before they began to write. Figures 2c and 2d represent writers who alternated often between planning and writing.

Correlations with quality. Although there was wide variation in the total time subjects spent writing, there was no relationship between the total time spent and overall quality, r = -.02. Subjects who spent more time writing did not necessarily writer better reports.

Nor did writers who planned before they wrote (as many composition teachers advise) write better reports. Writers who distributed their planning time throughout the session were just as likely to write good reports as those who concentrated their planning at the beginning of the session. In fact, the correlation between the overall quality and the Stage Index (which indicates the extent to which participants planned before they wrote) was slightly negative, r = -.17.

There was, however, a significant *negative* relationship between the quality of the reports and time spent planning. Writers who spent *less* time planning tended to write better reports. Table 3 shows the correlations between overall quality and several measures of the distribution of time. Each entry in Table 3 is a correlation between overall quality and a particular time measure. For example, the first entry, -.44, is the correlation between overall quality and time spent planning. Both total time spent planning and percent of time spent planning were negatively correlated with overall quality. The same general pattern held for both Network and Tree Modes.

Insert Table 3 about here

People who wrote lower quality reports not only spent more time in the planning modes, but did more work there. In order to measure the amount of work done, we counted number of "create" operations (operations that added to the structure), number of "change" operations (operations that changed the structure), and total planning operations carried out in Network and Tree Modes. Number of create operations and total number of planning operations were both negatively related to quality, r = -.39 and -.33, $p < .10$.

In light of the many claims that planning is good and that good writers plan more, the negative relationship between planning time and quality was very surprising. We had expected to find that those writers who spent more time planning would write better reports. How can these results be explained? This single study does not yield definitive answers, but we can make several observations.

The first is that many previous studies of planning involve high school or college students who, as a group, do little or no planning on their own. Many of these studies show that *some* planning is better than *no* planning. For example, Kellogg's (1987, in press) studies showed that college students wrote better text if they spent a few minutes planning than if they started

Table 3.

Correlations between overall quality and distribution of time between planning and writing/revising among the four system modes.

	Plan	Write	Network	Tree	Edit	Text
Time	-.44*	.21	-.28	-.45*	.09	.04
Percent time	-.48*	.48*	-.31	-.44*	.05	.14
No. of episodes	.05	.04	.12	-.02	.04	.29
Time/episode	-.39	-.14	-.23	-.42*	-.13	-.20

*p < .05

writing immediately. Just because some planning is better than none, we cannot conclude that more is always better. Rather, it seems likely, especially with respect to short texts, that there is some optimal amount of planning, and that time spent beyond that amount yields diminishing returns. (Note from Table 2 that the minimum amount of time our subjects spent planning was 22 minutes.) The fact that many of the reports produced in this study were overloaded with headings and that many headings had no associated text at all suggests that a number of our subjects spent too much time planning.

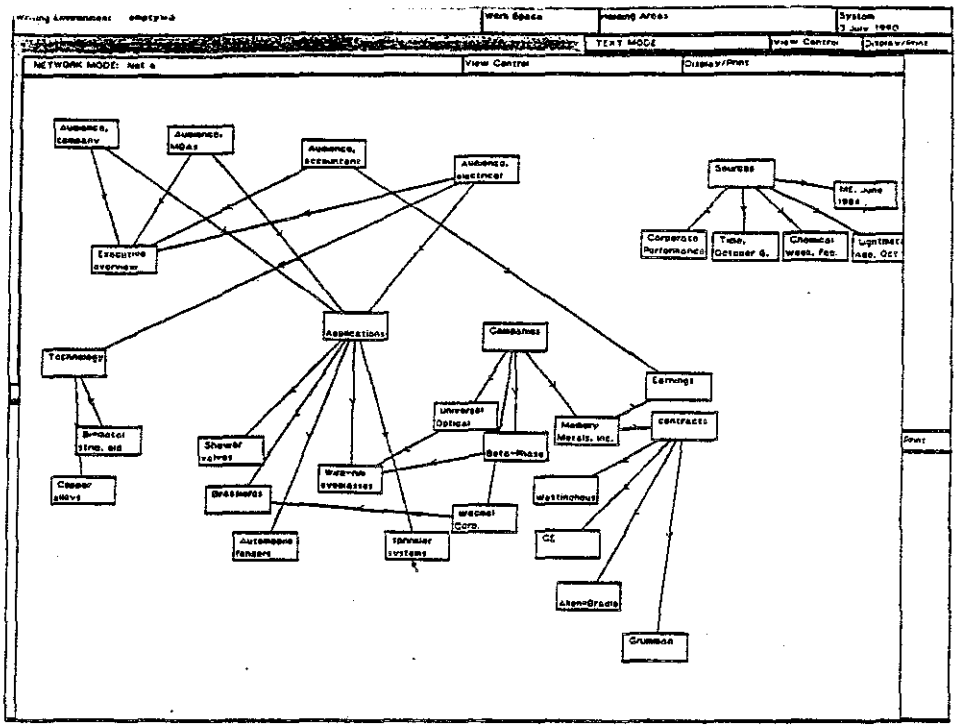
Replay of the writing sessions casts more light on these speculations. Some of our subjects appear to have spent more time experimenting with the novel capabilities of the Writing Environment, especially Network and Tree Modes, than was practical, given that they were required to complete their reports in a single writing session. Figure 3, which shows the organizational structure of one report at two stages in its development, illustrates an extreme case. This writer experimented with the idea of using nodes to represent individual readers of the report. (The experimental instructions specified that the report was intended for a management team including two MBAs, an accountant, an electrical engineer, and the owner.) In the end, he gave up the idea and reverted to a more standard organizational structure. But the experimentation had taken precious time away from other activities.

Insert Figure 3 about here

We can draw an analogy between previous studies of word processors and this study of the Writing Environment. Word processors increase the power of the writer to manipulate the wording of individual sentences. In response, writers seem tempted to spend too much time revising individual sentences. The planning modes of the Writing Environment increase the writer's power to define and manipulate the global structure of the text. In response, some of our writers seem to have spent too much time building and modifying their structures.

There seems to be a tendency for people to spend more time doing what the tool makes easy, especially when they are novice users. This does not mean that the tools should be discarded. People's attachment to their word processors indicates that they want the additional power provided by the computer. But it does mean that as they learn to use the new capabilities of the computer, writers must learn to let their writing strategies be dictated by their writing goals rather than by the capabilities of the system.

a.



b.

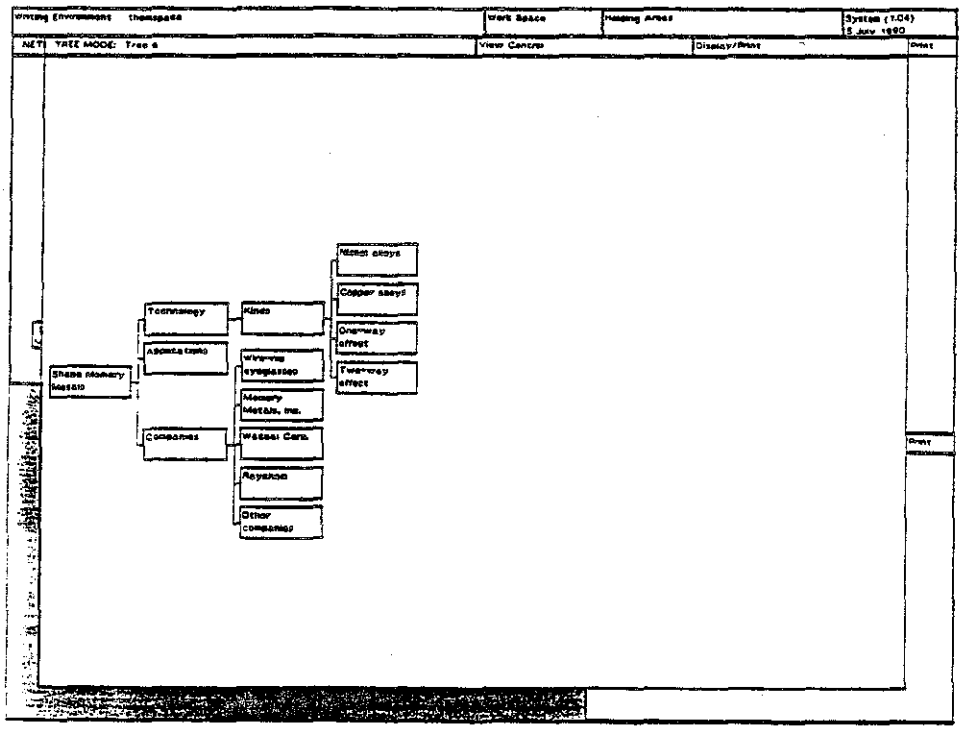


Figure 3. The two panels of the figure represent two phases in the development of a single subject's organizational structure. In Figure 3a, the subject was working in Network Mode experimenting with the idea of using nodes to represent the readers as well as the content of the report. Figure 3b shows the final structure of the report in Tree Mode. The structure has become much simpler, and the "reader nodes" of Figure 3a have been omitted.

We must caution that our subjects had only 1.5 hours of practice using the Writing Environment. As we do studies with more experienced users, we may find that the negative relationship between planning time and quality disappears or is reversed. In a more recent study, in which each subject wrote two reports using the Writing Environment, amount of time spent planning and number of nodes in the final structure both decreased from the first to the second report.

Finally, we must consider the validity of the judges' evaluations. Although we asked the judges to treat the reports as rough drafts and to weight content more heavily than style, they may still have been overly influenced by sentence structure as opposed to more global characteristics of the reports. In further research, we hope to use reader comprehension as an alternative measure of report quality.

Some Comments on Computer-Generated Protocols

Writing research falls roughly into two categories: studies that examine the thought processes of individual writers and studies that use standard statistical techniques to generalize over groups of writers (Strong, 1985). Representative of the first category is Emig's (1971) study of eight gifted high school writers, in which the strategies of only one student were given extensive, detailed analysis. Representative of the second category are Kellogg's (1987, in press) randomized controlled studies of college students who were instructed either to outline or to begin writing immediately.

While case studies may give the researcher a good feel for the conscious cognitive processes of individual writers, these studies are difficult to generalize. Analysis of think-aloud protocols is so labor intensive that researchers rarely report detailed data from more than a very few subjects. Although the informal insights derived from these protocols have changed the shape of the field, it is impossible to judge the validity of their more objective claims, such as the claim that expert writers elaborate the rhetorical problem more extensively than novices.

On the other hand, quantitative studies, such as those of Kellogg, often leave the reader wondering whether, in averaging over groups of subjects, the researcher has lost track of the often idiosyncratic nature of individual writers' strategies.

Computer-generated protocols, such as those produced by the Writing Environment, offer a way to bridge the gap between the case-study and the quantitative approaches. These protocols support both detailed study of individual writers' strategies *and* quantitative analyses of groups of writers. Watching the replay of individual writing sessions has given us a sense of the struggles individual writers go through as they try to shape their ideas into coherent organizational structures. Replay of individual sessions also reveals episodes during which the writer was struggling not

with ideas, but with the computer system, e.g., trying and failing to resize the screen so that both the organizational tree and the text are visible. Through replays we gain an intimate view of the subject, not only as a writer, but as a computer user.

But computer-generated protocols also tell the other side of the story: they allow us to summarize the data from groups of subjects and thus to judge the generalizability of our observations. Since the protocol is readable by the computer, we have been able to write computer programs that condense a detailed list of user actions into a higher level summary of user operations, classify and count operations, calculate the time spent on various types of operations, and summarize all these types of data over groups of subjects.

In this particular study, replay of individual writing sessions and quantitative analysis of summary data led to a coherent picture of user strategies. We discovered that writers had indeed learned to use the Writing Environment during the brief training session, and that, given the task of summarizing a set of source materials, they used the planning modes of the system extensively to organize those reports. In fact both replays of individual writing sessions and quantitative summaries indicated that at least some writers devoted too much of their effort to planning the organizational structures of their reports and too little to writing and revising the text.

Only further studies will reveal whether the findings reported here will generalize to experienced users writing longer texts.

References

Berkenkotter, C. (1983). Decisions and revisions: The planning strategies of a publishing writer. *College composition and communication, 34*, 156-169.

Bereiter, C. & Scardamalia, M. (1987). *The psychology of written composition.* Hillsdale, NJ: Erlbaum.

Bridwell-Bowles, L., Johnson, P., & Brehe, S. (1987). Composing and computers: Case studies of experienced writers. In A. Matsuhashi (Ed.), *Writing in real time.* Norwood, NJ: Ablex.

Collier, R.M. (1983). The word processor and revision strategies. *College Composition and Communication, 34*, 149-155.

Daiute, C. (1986). Physical and cognitive factors in revising: Insights from studies with computers. *Research in the Teaching of English, 20*, 141-159.

Emig, J. (1971). *The composing processes of twelfth graders.* Urbana: National Council of Teachers of English Research Report No. 13.

Faigley, L., Cherry, R.D., Jolliffe, D.A., & Skinner, A.M. (1985). *Assessing writers' knowledge and processes of composing.* Norwood, NJ: Ablex.

Flower, L. & Hayes, J.R. (1980). The cognition of discovery: Defining a rhetorical problem. *College Composition and Communication, 331*, 21-32.

Flower, L. & Hayes, J.R. (1981a). A cognitive process theory of writing. *College Composition and Communication, 32*, 365-387.

Flower, L. & Hayes, J.R. (1981b). Plans that guide the composing process. In C.H. Frederiksen & J.F. Dominic (Eds.), *Writing: The nature, development and teaching of written communication.* Hillsdale, NJ: Erlbaum. pp 39-58.

Gould, J.D. (1980). Experiments on composing letters: Some facts, some myths, and some observations. In L.W. Gregg & E.R. Steinberg (Eds.), *Cognitive processes in writing.* Hillsdale, NJ: Erlbaum.

Gould, J.D. (1981). Composing letters with computer-based text editors. *Human Factors, 23*, 593-606.

Haas, C. (1989). Does the medium make a difference? Two studies of writing with pen and paper and with computers. *Human-Computer Interaction, 4,* 149-169.

Haas, C. & Hayes, J.R. (1986). What did I just say? Reading problems in writing with the machine. *Research in the Teaching of English, 20*, 22-35.

Hagge, J. (1987). The process religion and business communication. *Journal of Business Communication, 24*, 89-120.

Hairston, M. (1982). The winds of change: Thomas Kuhn and the revolution in the teaching of writing. *College Composition and Communication, 33*, 76-88.

Hawisher, G.E. (1987). The effects of word processing on the revision strategies of college freshmen. *Research in the Teaching of English, 21*:2, 145-159.

Hayes, J.R. (1989). Writing research: The analysis of a very complex task. In B. Simon, D. Klahr & K. Kotovsky (Eds.), *Complex information processing: The impact of Herbert Simon*. Hillsdale, NJ: Erlbaum.

Jenkins, I., Lansman, M., & Smith, J.B. (1989). *A tutorial for WE 1.0 for use with experimental studies*. Chapel Hill, NC: UNC Department of Computer Science Technical Report # TR89-016.

Kellogg, R. T. (1986). Writing method and productivity of science and engineering faculty. *Research in Higher Education, 25*, 147-163.

Kellogg, R.T. (1987). Writing performance: Effects of cognitive strategies. *Written Communication, 4*, 269-298.

Kellogg, R.T. (in press). The effectiveness of prewriting strategies as a function of task demands. *American Journal of Psychology*.

Kellogg, R.T. & Mueller, S. (1989). *Cognitive tools and thinking performance: The case of word processors and writing.* Paper presented at the annual meeting of the Psychonomic Society, Atlanta, November, 1989.

Lutz, J.A. (1987). A study of professional and experienced writers revising and editing at the computer and with pen and paper. *Research in the Teaching of English, 21*, 398-421

Mischel, T. (1974). A case study of a twelfth-grade writer. *Research in the Teaching of English, 8,* 303-314.

Pianko, S. (1979). A description of the composing processes of college freshman writers. *Research in the Teaching of English, 13*, 5-22.

Selzer, J. (1983). The composing processes of an engineer. *College Compositions and Communication, 34*, 178-187.

Smith, J.B. & Lansman, M. (1989). A cognitive basis for a computer writing environment. In B.K. Britton & S.M. Glynn (Eds.), *Computer writing environments* (pp. 17-56). Hillsdale, NJ: Erlbaum.

Smith, J., Rooks, M.C., & Ferguson, G.J. (1989). *A cognitive grammar for writing: Version 1.0.* Chapel Hill, NC: UNC Department of Computer Science Technical Report TR 89-011.

Smith, J.B., Weiss, S.F., Ferguson, G.J., Bolter, J.D., Lansman, M., & Beard, D.V. (1987). WE: A writing environment for professionals. *Proceedings of the National Computer Conference '87* (pp. 725-736). Reston, VA: AFIPS Press.

Stallard, C.K. (1974). An analysis of the writing behavior of good student writers. *Research in the Teaching of English, 8,* 206-218.

Stotsky, S. (1990). On planning and writing plans - or beware of borrowed theories! *College Composition and Communication, 41*, 37-57.

Strong, W. (1985). Linguistics and writing. In B.W. McClelland & T.R. Donovan (Eds.), *Perspectives on research and scholarship in composition.* New York: Modern Language Association.

Swarts, H., Flower, L.S., & Hayes, J.R. (1984). Designing protocol studies of the writing process: An introduction. In R. Beach and L.S. Bridwell (Eds.), *New directions in composition research.* New York: Guilford.

Author Notes

This research was supported by Army Research Institute Contract #MDA 903-86-C-0345 and the National Science Foundation Grant #IRI-8519517. We would like to thank Gordon J. Ferguson, who was the principal technical designer of the Writing Environment and provided assistance with the system throughout the experiment and data analysis. We are also grateful to Matthew Barkley, Paulette Bush, Karen Harrower, Mark Rooks, and Yen-Ping Shan for technical support, and Gregg Berg for assembling materials and instructions.

The first author was a visitor in the Department of Technical Communication while writing this paper.

Requests for reprints should be sent to Marcy Lansman, Department of Computer Science, CB #3175, University of North Carolina, Chapel Hill, NC 27599-3175.