Reading and Writing with Computers:
A Framework for Explaining
Differences in Performance

Wilfred J. Hansen, Information Technology Center

Christina Haas, Information Technology Center
and English Department

Carnegie-Mellon University
4910 Forbes Ave.
Pittsburgh, PA 15213

Keywords: Computer-Human Interaction, Text Editing, User Interface, Reading and Writing with Computers, Human Factors, Word Processing, Andrew

Summary: Reading from computer screens is increasingly important as a source of timely information. Writing with computers is increasingly popular for its rapidity and ease of revision. Since numerous studies have shown that reading and writing are distinctly different with computers and paper, it is now time to ask whether there is an overall explanation of the differences and why they occur. In this article we answer these questions by describing seven factors that influence reading and writing with computers: Page Size, Legibility, Responsiveness, Tangibility, Sense of Directness, Sense of Engagement, and Sense of Text. These factors are illustrated by showing how they can explain the results of a series of experiments we conducted.

1. Introduction

Reading and writing with computers are increasingly important tasks as the volume of information in machine-readable form increases. For faster access, library material itself--in addition to indices--is being made available through computers. The EXPRES project [NSF, 1986] is automating proposal submission to the NSF so principal investigators, reviewers, and administrators can all access the proposal via computer; conceivably they might never be read from paper at all. Material from newspapers to encyclopedias is being made available on-line. The instruction manuals to aid users are themselves readable from the screen.

A number of studies, including our own, have been conducted of user behavior while reading and writing on-line [Gould, 1981; Haas and Hayes, 1985a, 1985b, 1986a, 1986b; Gould and Grischkowsky, 1984; Hawisher, 1987]. However, due to the complexity of the cognitive tasks involved, and the variety of experimental paradigms, it has been difficult to interpret and compare results across studies. It is our purpose in this paper to present a framework of factors within which the variations among results can be explained. After a review of the literature in this section, we will describe the seven factors and then show how these factors explain our results. This paper is not a report of our experiments; all have been reported in detail elsewhere as cited in the individual sections below.

Most studies have found that reading from paper is faster than reading from computer screens. Muter, et al. [1982] showed that reading from TV screens took 25% longer than from paper, but produced roughly equal comprehension scores. Wright and Lickorish [1983] also found that paper was faster. Gould and Grischkowsky [1984] studied subjects performing an eight hour proof reading task. They found that work was more rapid on paper, with slightly higher quality than on personal computers. Our own experiments verified these results and extended them to positional memory and various alternate computer conditions.

Results are more contradictory for writing tasks. Gould [1981] found that expert writers using personal computers required 50% more time to compose than on paper, while producing texts judged to be of no greater quality. Hansen, Doring, and Whitlock [1978] showed that students took considerably longer to answer an examination on-line rather than on paper, though a large portion of the difference could be attributed to poor design of the interactive interface. Our results were consistent with these for writing with personal computers, but strikingly different with advanced workstations of the class typified by the IBM RT/PC with a large-screen bit-mapped display, large memory, high speed processor, and a mouse. On the latter, subjects spent more time on the task, but the number of words per minute was the same and the quality of text was actually superior.

The environment for our work was the development of the Andrew system [Morris, et al, 1986]. It was our hope and that of other system designers that a better system could be deployed if we paid careful attention to the user interface, including the conduct of controlled experiments to explore alternatives. At the same time, one of the co-authors (Haas) was exploring paper versus computer as a medium for reading and writing. These studies seemed an ideal vehicle for exploring the emerging user interface for the Andrew text editor, Edittext.

Our experiments utilized among them five different media conditions: one with paper and pencil, two with personal computer, and two with personal workstation. A paper version of each experimental task served as a control; otherwise each experiment employed only one or two of the computer conditions.

Of the two personal computer conditions, one used it as a terminal and the other as a local computer. As a terminal, it ran Emacs [Stallman, 1981] on a mainframe computer (TOPS-20), connected at 4800 or 9600 baud. As a local computer, subjects had a choice of two editors--Mince and Epsilon--both similar to Emacs. The two workstation conditions utilized Edittext on Andrew, varying the size of the window between large and small. See Figure 1 for examples of both screen size conditions.

[Insert Figures 1(a) and 1(b) here.]

---

2. Factors

Our experiments and observations can be explained with seven factors--four primary and three secondary. These factors are not all original with us, nor are they the outcome of a factor analysis or other statistical process. It is unlikely that these factors alone account for the observed effects. Despite these limitations, the factors provide a convenient framework to organize our results and discuss the multitude of influences at work when people read and write using computers.

Primary Factors

The primary factors are directly observable attributes of hardware and software design, variations of which may affect performance. Each is a distinct dimension which can be varied independently in further experiments. The four primary factors are Page Size, Legibility, Responsiveness, and Tangibility.

A. Page Size is the amount of text visible at one time. It can affect reading and review tasks by limiting the context for the visible text, thus burdening short-term memory. It can affect writing by impeding reference to recently written text, possibly leading to repetition or omission. If the Page Size is small, the user will have to scroll more often to view the entire text. Not only do these take time, but each interferes with concentration. For example, one study estimated that there was a three second pause for a subject to re-establish contact with the work when the screen was repainted [Hansen, 1978].

Our experiments utilized two page sizes, Small and Large. The archetype of the Small size is the screen on the personal computer, holding 24 lines of 80 characters each. The small window condition on the personal workstation was adjusted to hold about the same number of characters; it utilized a space 5 1/2 inches high by eight inches wide holding 22 lines of variable-width text. This space held only about forty percent of the contents of a sheet of paper. The Large page size condition--on workstations only--displayed a full page of text. The window was approximately 10 inches square and held 46 lines of about 80 characters each.

B. Legibility is the ease with which letters and words can be correctly recognized. That legibility has a strong influence on reading speed has been reported in Gould [1987] and Booth [1987]. While many characteristics contribute to legibility: font design, spacing, contrast, edge sharpness, anti-aliasing, flicker, resolution, . . ., none is pre-eminent. As Gould points out "each variable contribut[es] . . . in a small, cumulative way."

Since so many factors contribute, it is difficult to objectively measure legibility. We judged that the paper forms of our experiments had higher legibility than the computer versions because the resolution and edge sharpness were higher. The greater resolution also permitted use of standard fonts. The workstation conditions offered higher quality text than the personal computers: the workstation had a black-on-white image, proportionally spaced and seriffed fonts, and headings in boldface, larger type, or both. Resolution was 72 pixels to the inch. Characters on the personal computer had a resolution is 70 pixels per inch vertically and 80 horizontally. One advantage for the personal computer may be the greater contrast of the green on black; green is near the optimum wavelength for visibility.

C. Responsiveness is the speed of system response to a user's action and has two components: the speed with which the system begins to respond and the speed with which it completes its response. Typically the response to a text key is an instantaneous display of the character. The response to a scroll request begins immediately, but may take one or more seconds to complete. The response to a *print* command may take minutes as the document is formatted for the printer.

The psychological impact of a slower response can depend on the user's state of *completion* as the action is performed. Completion is a measure of the degree to which the user feels finished with a phase of an operation. Typing a text key has low completion because the user is concentrating on text to come. Printing a document usually has higher completion, because user is committing the work to paper and has therefore probably finished a phase of the creative work. Scrolling operations generally have low completion because the user is anticipating new information and can do nothing until it appears. Poor Responsiveness when the user has a low degree of completion can be frustrating and may induce errors. Thus slow system response can delay a user not only by causing operations to take longer, but also by reducing concentration and making errors more likely.

The Responsiveness for moving through a document is excellent with paper, although its Responsiveness for writing may be low, especially for children. With a personal computer, the Responsiveness when used for local editing is generally good, depending on the editor in use. As a terminal, the personal computer is no better than the host system and is limited by the speed of the communication line. At 4800 baud, the repaint time for a screen-full is two seconds. In contrast, the workstations using Andrew required less than a second to repaint even the large window.

D. Tangibility describes the extent to which the state of the system appears to the user to be visible and modifiable via physical apparatus. An intangible representation of a numeric value might be a sequence of digits; a more tangible representation of the same value would be a dial. The system is even more tangible if the value can be adjusted by clicking on the dial or dragging the needle (such designs have been called "direct manipulation" by Shneiderman [1983]). Even data base entries displayed with images of spiral bindings or slotted index cards enhance the impression that the computer is presenting tangible facts rather than ephemera.

Tangible designs are important, we believe, because they aid in learning, remembering and efficiently using a system. A pictorial representation is easier and faster to comprehend than textual information and modification of images often avoids the need to design, document, teach, and support a plethora of commands. We have not directly tested these assertions, but there are consistent with the results of our experiments, wherein more tangible systems enabled better performance, at least by the non-expert user.

A valuable tool in the design of a tangible system is a mouse or other pointing device. Without such a device, pointing at positions on the screen can only be done rather indirectly by typing a sequence of keys. Note that we are only claiming here that the mouse aids Tangibility, not that it is superior to keystroke sequences. A very real question, whether manipulation of Tangibile graphic images is superior to keystroke sequences, remains unanswered.

Text on paper has high Tangibility: it is laid out in particular places on each sheet of paper, the sheets are stacked together, and the user can move sheets from the unread stack to the finished stack. As the user reads, the shifting stack gives gives tactile position cues as well. This contrasts with the editors used on mainframes and personal computers in our experiments. At best the text of each file is accompanied by message line in which an integer indicates the position within the document of the visible image. The only mechanism available to view another portion of the document is keystroke sequences.

The viewing of text with Andrew is somewhat more Tangible. A *scrollbar* displays an analogue representation indicating which portion of the document is visible. The scrollbar is a vertical rectangle at the left of the text which represents the entire length of the text. An *elevator* image within the scrollbar displays graphically both the position and extent of the visible text. Mouse operations within the scrollbar can change the view to an adjacent or remote part of the text. When a mouse click scrolls to the next page, a few lines from the bottom of the previous page are left at the top of the screen to provide continuity. Thus in discussion of our experiments the number of scrolling operations required to move through an entire document are larger than the number of screen-fulls required to display the document.

Another Tangibility aspect of the systems employed in our experiments was the selection of an area of the text and the indication of what is selected. With paper, a section can be selected by physically boxing or bracketing an area. With the various personal computer conditions, the cursor can be moved and selections can be marked via keystrokes; in some cases the selected text is highlighted. With Andrew a section of text is selected by pointing at its ends with the mouse and clicking. In our experiments, a box was drawn around the selected text.

Secondary factors

Rather than claim that the primary factors are immediately responsible for user behavior--and thus for the results of our experiments--we posit a set of secondary factors. Each secondary factor is itself determined by a combination of the primary factors and induces a state or "Sense" within the subject. Possible interactions of primary and secondary factors and their possible influences on user performance are diagrammed in Figure 2.

[Insert Figure 2 here.]

A. Sense of Directness. A user's Sense of Directness is the degree of feeling that the changes are the screen are a direct result of the user's actions. Ideally, the user has an *illusion of mechanical linkage*, a feeling that the displayed image is a physical object which the user can manipulate as easily as turning the pages of a book or writing a note in the margin.

A Sense of Directness helps a user learn and internalize the interface to a system because each response by the system reinforces the user's confidence and understanding. Directness changes the way a user interacts. With an indirect system the user thinks about a problem, decides on a change, enters a command, observes the response, and repeats. With a Direct system the user should be able to think about a problem and make a change, without thinking about how the change is made. The interaction is so natural that the user ceases to think about it, just as a manual writer seldom pays any attention to paper and pencil. In terms of Hansen [1971], the user utilizes "muscle memory" rather than conscious control.

The Sense of Directness is strongly affected by the Responsiveness and Tangibility of the system. Since paper is high on both these factors, it generally engenders a high Directness. The Andrew conditions are second highest in Responsiveness and Tangibility, so should engender a Sense of Directness not far below that of paper. That the personal computer conditions should have lower Directness is reasonable since the system designs are considerably less Tangible; this is not offset by Responsiveness, which is little better when the personal computer is used by itself and is considerably worse when used as a terminal.

B. Sense of Engagement is a feeling that the system is holding an interesting, and even fascinating, conversation with the user. At its extreme, it induces a state of intense, almost addictive concentration similar to an exciting two-person game. One source of Engagement with systems is the fun of seeing the system react; another is similar to the fascination exhibited by subjects in stimulus-response experiments. The instant response of the computer provides a reward which reinforces the user's behavior.

A good interactive system harnesses Engagement to keep the user interested in his or her task for longer periods than other systems might. However, Engagement may not always be a desirable response to an editor. Quality of work may decline if too much time is spent on task; indeed, the ease of making local changes to a text with word processing may distract writers from attending to other important writing concerns [Haas, 1987].

One facilitating factor inducing Engagement, especially in novices, is Tangibility. It is possible to have intense communication via more abstract interaction with the keyboard, but a level of skill must be reached before this takes over from the confusions of trying to learn how to use the system.

A more important factor is Responsiveness. Fast reaction encourages the user to respond rapidly in turn, setting up a rhythm of intense interaction, while slow response gives the user time to be distracted and lose concentration. Systems with variable Responsiveness, perhaps due to multi-processing, not only interfere with concentration, but may even cause frustration. Possibly they are harder to learn to use, just as subjects in stimulus-response experiments exhibit longer learning times when treated with variable reinforcement schedules.

For our experiments we consider that paper had low Engagement because it is familiar to users and non-interactive. Personal computers used as terminals probably had negative Engagement because of slow and variable response. Personal computers for local editing can have very good Engagement because the response can be instantaneous and invariant. The workstation editor is not yet quite Responsive enough to have Engagement as high as the personal computer when the keyboard alone is used. However, use of the mouse to position the cursor seem to have a novelty and Directness that generate considerable Engagement.

C. Sense of Text. One difficulty users have dealing with documents on computers is in getting a *sense of the text* [Haas and Hayes, 1985]. By this phrase we mean the feeling that a user may have that he or she has a good grasp of the structural and semantic arrangement of the text--the absolute and relative location of each topic and the amount of space devoted to each. Good Sense of Text is invaluable to a reader in finding parts of the text, following the thread of an argument, and forming a "gist" of the material. For a writer, Sense of Text has all these merits and is necessary in order to organize the text effectively, avoid duplication, and assess whether plans and goals have been met.

Rothkopf [1971] has shown that readers can recall the position of text on paper pages. This may aid Sense of Text by tying the text to a physical entity which provides visual and tactile cues. Furtherwe have found that writers may spend less time planning when writing with computers than when writing with paper [Haas, 1987]. Presumably, time spent planning the text may be partially spent rehearsing its structural and semantic content. Writers may have a problem "getting a sense" of their computer-produced texts because they spent less time planning them.

Many factors may detract from a Sense of Text with computers. The position of lines within pages cannot be known if the computer system displays text with a different line at the top of the window each time. A small Page Size reduces the context for each piece of text. Limited Legibility may cause the reader to spend more mental effort on recognizing individual words and comparatively less on getting an impression of the entire page. Even poor Responsiveness may distract the reader with delays while scrolling. However, the Sense of Text could be enhanced by the Tangibility of a scrollbar.

Since the other six factors--Legibility, Page Size, Responsiveness, Tangibility, Sense of Directness, and Sense of Engagement--seem likely to impact the reader’s Sense of Text, we chose to study this factor with our first three experiments.

3. Experiments and Results

In this section we review four experiments we have conducted to study various aspects of the factors affecting use of computer for reading and writing. The results are summarized in Table 1. The left-hand columns compare the various computer conditions with respect to the primary factors, treating paper as the norm, while the right-hand columns compare the experiment results. An asterisk (*) indicates a result that differs with statistical significance from other results on that experiment. Note that significant differences in results occurred when Responsiveness was greatly inferior or when some other condition was inferior.

[Insert Table 1 here.]

The first three experiments required subjects only to read material. Responses were given verbally or by pointing with a finger. Interaction with the computer was limited to scrolling the text, which used keystrokes on the keyboard or mouse clicks in the scrollbar. We believe the important factors in the results of these reading experiments are Page Size, Legibility, Responsiveness, and Sense of Text.

A. Spatial Recall

Spatial recall is the ability to remember the page and line of specific items. Rothkopf [1971] found that subjects reading from printed text showed significant spatial recall. Since this ability may be an important component of Sense of Text--allowing readers to remember the location and arrangement of points of a text--this experiment was designed to study how spatial recall is affected by viewing a text on a computer screen or on paper. [For full details of this study, see Haas and Hayes, 1985a; 1986a.]

The subjects were familiar with the text editor used; five subjects performed the task on paper and five on the personal computer used as a terminal. Subjects read a text of 1000 words (nine pages or screens) and were subsequently shown eight particular sentences from that text and asked to mark their location on a blank image of the text (empty paper in a folder or blank lines in a text file). The text presented on each paper page was the same size as text presented on line: there were nine screens/pages in each condition. The responses were compared with the correct page, line, and position in line and the scores assigned as the difference between the response and the correct answer in each category.

Results showed that subjects' responses were more accurate when they read from paper. Significant differences were found for the line-on-page (vertical) recall variable: mean differences between recalled location and actual location (for eight trials) was 30 lines in the paper conditions and 45 in the computer condition. This difference was significant (p<.05) by analysis of variance. A larger effect for this variable is not surprising since vertical location is not consistent with a scrolling screen.

Of the four primary factors, Page Size cannot explain the observed differences because pages were the same size in both conditions. The difference in Responsiveness was large--two seconds per page on the computer--and may have been a major cause of the performance differences. However, the computer also had lower Legibility and lacked the rudimentary Tangibility afforded by the thickness of paper as pages are turned from one pile to another. In any case, the subjects' Sense of Text (as measured by spatial recall scores) seems to have been impaired by the computer condition.

B. Information Retrieval

The first experiment demonstrated that readers can recall the location of information more accurately from paper than from a personal computer. This result suggests that readers would find it easier to retrieve information to answer questions from paper than from computer screen. The second experiment was designed to test this possibility. There were three conditions: paper, the advanced workstation with Andrew, and the personal computer used as a terminal to a mainframe.

After reading an 1800-word text, subjects were asked to retrieve answers from the text to a series of questions. The paper version of the experiment was printed in twelve point TimesRoman, the personal computer version utilized the green monochrome display, and the workstation version was the large screen condition with twelve point TimesRoman text, but with bold text to highlight headings instead of all-capitals as used in the other two conditions. As formatted, the text occupied 3 1/2 pages on paper, 12 scroll operations on the personal computer, and 5 1/2 scroll operations on the workstation. Subjects were students familiar with using the personal computer as a terminal, although unfamiliar with Andrew. The subjects in the Andrew condition received training before the experiment on the use of the mouse pointer and scroll bar, which they used to move through the document. Facility with other Andrew features was not necessary for this experiment. [For full details of this study, see Haas and Hayes, 1985a; 1986a.]

Condition was a between-subjects variable; i.e., each subject did the experiment in only one condition. Almost all responses were correct so the performance measure was not accuracy, but total time to complete the retrieval task. The mean time to complete the task was greatest in the personal computer conditions, 32.7 minutes; mean time for the workstation condition (with a large window) was 15.9 minutes; and for the paper condition, 13.0 minutes. The differences between conditions were significant ($p<.05$) by analysis of variance; Neuman-Keuls analysis revealed significant differences ($p<.05$) between the personal computer condition and the other two conditions, which did not differ from one another.

It is not surprising that there is a large difference in performance between the two systems: most of primary factors outlined earlier differed in this experiment. The Page Size differed by a factor of more than two; the Legibility of the workstation text was enhanced by a seriffed font and bold headings; Andrew was more Responsive both in beginning to respond to a command and in displaying a page; and Andrew utilized the Tangibility of the scrollbar for moving through the document. It seems probable that the performance differences indicate differences in Sense of Text; if so, we argue that the primary factors strongly influence this Sense.

C. Reordering a Scrambled Text

There are any number of variables which may account for the results of the information retrieval experiment. We hypothesized that the size of the screen could be s significant factor in both the results of that experiment and in readers' Sense of Text. Experiment C, Reordering a Scrambled Text, was designed to isolate two factors--page size and tangibility--and assess their impact on subjects' performance. The experiment crossed size variables (large and small windows) and methods of scrolling (scroll bar and function keys). Again, a paper condition served as a control. A within-subjects design was used for this experiment, with the order of the conditions counter-balanced. In a within-subjects design, each subject serves as a control for his or her own performance, thus eliminating the impact of individual characteristics like reading and typing speeds.

The experimental task tested the ability of subjects to read critically in order to determine the correct arrangement of a disordered text. Critical reading requires forming a mental representation of a text's content and is a more sophisticated skill than Spatial Recall or Content Retrieval. This kind of reading is necessary when revising or reorganizing a text and requires an understanding of the whole text, rather than just local interpretation.

In each condition of the experiment, subjects read a 1200 word text whose lines were scrambled and numbered. The texts used for the experiment were all taken from freshmen-level textbooks and were of similar readability. The subjects, all incoming freshmen at Carnegie-Mellon, were each given three hours of individual training on the workstation to become familiar with the system and the two scrolling methods. To reduce interference from motor variables, subjects responded orally; they gave instructions (by line number) as to how the text should be re-sequenced to produce a meaningful whole. [For full details of this study, see Haas and Hayes, 1985b; 1986a.]

Subjects performed the task in five conditions: paper and four workstation conditions which crossed the variables of window size and method of scrolling. On paper and with the large window, the texts occupied about two pages; with the small window the texts were about 4 1/2 pages. Two methods for moving the text were tested: one method was the Andrew scrollbar and the other, four function keys: page forward, page back, beginning of document, and end of document.

The subjects' error rates were low and uniform, so the results, shown in Table 2, are displayed as the mean time to complete the task. Subjects did best on paper, less well with large windows, and poorly with the small window. The differences between large and small windows, and between paper and small windows were significant; the difference between paper and large windows was not. Method of moving through the text--scroll bar or function keys--made no significant difference.

[Insert Table 2 here.]

In this experiment, Page Size seems to be an important factor: the task of rearranging lines was made more difficult in the small window because subjects had to scroll back and forth to understand the relations among the lines. Legibility was identical for all computer conditions. While Response time was identical for the computer conditions, the larger number of scroll operation for the small windows may have increased the total time slightly. It is unlikely that time for scrolling operations would not have accounted for all the difference between large and small windows, however, because a scroll operation in the small window only takes about half a second. The additional time with the small windows may be due to decreased Sense of Text in that condition.

The scroll bar increased the Tangibility, but this experiment revealed no difference in performance between scroll bar and function keys. In a separate experiment studying proofreading [Haas and Hayes, 1985b], subjects were allowed to choose between function keys and scrollbar; almost universally they chose the scrollbar. They may have preferred its Tangibility.

D. Letter Writing

Our fourth study examined writing, a task which may require more interactive behavior than the reading tasks examined in the previous experiments. Writing is a particularly difficult task because the author must create, review, and revise the text in light of purpose and intended readers. This very complexity makes it an interesting task to examine experimentally.

To study writing behavior, we chose as a paradigm Gould’s [1981] study comparing writers’ performance writing letters with text editors and paper. Fifteen experienced writers who regularly used computers like those on which they were tested, were asked to write a persuasive letter to a specific audience in three conditions: paper, a local editor on a personal computer, and the Andrew editor on a workstation. Each subject wrote in all three conditions, with topics and order counter-balanced. Gould's study had employed line editors; we hypothesized that hardware and software advances might lead to different results. Both large and small screen conditions were used, but there were no differences between subjects' performance with large and small windows, so the results for the two window sizes have been collapsed for this analysis. [For full details of this study, see Haas and Hayes, 1986b; and Haas, in preparation.]

Two quantitative measures were collected and analyzed: time to complete the task and number of words. Results are summarized in Table 3. In all conditions the words produced per minute were about the same, but subjects worked longer and wrote more words on the workstation. There were significant differences between the workstation condition and the other conditions in both time to complete the task and number of words produced. In these quantitative measures, subjects seemed to perform similarly with personal computers and with paper, and differently with the workstation.

The findings were slightly different for the qualitative measures, Content Quality and Mechanics Quality. For both, quality was evaluated by a forced quartile split. Two independent readers, with at least five years experience teaching English, rated each set of letters and were instructed to place each letter into one of four quartiles. In this way, each subject's letters were rated only against themselves. The quality score was the sum of the two quartile scores and ranged from 2 to 8. Agreement between the raters was about eighty percent, and a third rater was used for scores that differed by more than one quartile. Content Quality (ideas and supporting information) and Mechanics Quality (surface level correctness) were measured separately and later summed to produce a Total Quality score. Texts produced with the workstation and texts produced with pen and paper were significantly better than those produced with the personal computer both in Content Quality and in Total Quality.

Our results show a statistically significant differences in quality among letters while the results of Gould's earlier study did not. Our finding of a difference may have been due to the method of assessing quality: Gould's evaluators gave an independent grade to each letter, while ours used the more discriminating quartile split. We hypothesize that a forced quartile split may be more sensitive than Gould's measure of quality.

[Insert Table 3 here.]

The results of this experiment raise several questions:

Why the disagreement with Gould's results? Gould's subjects were 50% slower with computers while our subjects had the same or better speed on computer than on paper. The difference is probably because Gould's subjects used a line editor rather than a full-screen editor. The Page Size with a line editor is effectively smaller because the user must make a request in order to see more text. More importantly, Tangibility is poor: commands must be issued to make changes and they actually appear on the screen in the same areas as text.

Why did subjects work longer and produce more words with the workstation? Possibly subjects worked longer to fill a larger available Page Size; however, results were similar with both large and small windows. Perhaps the work was physically easier because typing is easier than handwriting; however, the personal computer shares the same physical ease. Our favorite explanation is that subjects felt more Engagement with the workstation and worked longer for more self-satisfaction.

Why was the quality of work higher on the workstation and with paper than on the personal computer? We may speculate as follows: Legibility and Page Size make it easier to review one's text the workstation and paper than on the personal computer. The Tangibility of the workstation reduces the number of commands that must be typed, reducing confusion with the text that must also be typed. In the paper condition, this confusion would not be present at all. Both the workstation and the paper are more Responsive and Direct and may encourage a heightened Sense of Text. All these factors may work together, reducing non-productive efforts and freeing the subject to think about carrying through a cogent argument.

4. Conclusion

In this paper we have sketched the factors affecting user performance when reading and writing with computers. We have described four primary factors--Page Size, Legibility, Responsiveness, and Tangibility--and three secondary factors--the Senses of Directness, Engagement, and Text. These factors were then used to explain differences observed in four experiments.

Every experiment showed that paper was superior for reading to any computer condition, although the workstation results were closer to those of paper than those of the personal computer. On the writing task, paper differed from the personal computer chiefly in that subjects produced higher quality letters. In addition, subjects worked longer and wrote more with workstation than with the other media.

It would not be fair to claim that workstations are universally superior to personal computers. With both available, one of the authors of this text prefers a personal computer because of the advantages of the Personal Editor [Wylie, 1982]. Its Responsiveness and the resulting feelings of Directness and Engagement outweigh the disadvantages of reduced Page Size and lack of Tangibility. However, there is a considerable feeling of loss of Sense of Text, which must be offset by producing a paper copy of the text for review and markup.

How have our seven factors helped explain the observed results? Table 1 summarizes the situation for the four primary factors: where a medium was very inferior to paper on at least one of the primary factors, we found a statistically significant deterioration in performance. For the secondary factrors, the first three experiments all showed effects that can be explained as loss of Sense of Text in the subjects. The fourth experiment--the only one with a writing task--showed effects that we explain as revealing differences in the Senses of Directness and Tangibility.

The final test of our work for the Andrew project must be whether the studies reported here had a favorable influence on the user interface finally deployed. In fact, they did. Numerous changes to the system were made in response to observations made during the conduct of these and other studies, including changes to the scroll bar and menus, establishing default window size and placement, and choosing fonts and margins for the text editor. It may well be that the most important result of user interface studies are not the findings of specific experiments, but fostering of a general attitude of adapting and modifying the computer system to the users it is intended to serve.

Acknowledgments: We are grateful for the help and guidance of John R. Hayes of the CMU Psychology Department who collaborated on the experiments reported here. Christine M. Neuwirth of the CMU English Department also advised on several of the experiments. James Gosling built the first version of Edittext, with contributions from W. J. Hansen and A. J. Palay. We are grateful for their help and for the help of all other members of the Information Technology Center and its director, James H. Morris.

References

Booth, K. S., Bryden, M. P., Cowan, W. B., Morgan, M. F., Plante, B. L. On the parameters of human visual performance: An investigation of the benefits of anti-aliasing. In *Proceedings of CHI+GI 1987* (Toronto, April 5-9). ACM, New York, 1987, pp. 13-20.

Gould, J. Composing Letters with Computer-Based Text Editors. *Human Factors 23*(5), 1981, 593-606.

Gould, J., and Grischkowsky, N. Doing the same work with hardcopy and with CRT terminals. *Human Factors 26*(3), 1984, 323-337.

Gould, J. D., Alfaro, L., Finn, R., Haupt, B., Minuto, A., Salaum, J. Why reading was slower from CRT displays than from paper. In *Proceedings of CHI+GI 1987* (Toronto, April 5-9). ACM, New York, 1987, pp. 7-12.

Haas, C. *How the Writing Medium Shapes the Writing Process: Studies of Writers Composing with Pen and Paper and with Word Processing.* Doctoral dissertation, Carnegie Mellon University, 1987.

Haas, C. Does the medium make a difference: a study of writers composing with pen and paper and with computers, in preparation.

Haas, C., and Hayes, J. *Effects of text display variables on reading tasks: computer screen vs. hard copy.* Pittsburgh: CDC Technical Report #3, 1985a.

Haas, C. and Hayes, J. *Reading on the computer: a comparison of standard and advanced computer display and hard copy*. Pittsburgh: CMU, CDC Technical Report #3, 1985b.

Haas, C. and Hayes, J. What did I just say? Reading problems in writing with the machine. *Research in the Teaching of English*, February, 1986a.

Haas, C. and Hayes, J. R. *Pen and paper vs the machine: Writers composing in hard copy and computer conditions.* Pittsburgh: CDC Technical Report #16, 1986b.

Hansen, W. J. User Engineering Principles for Interactive Systems, Fall Joint Computer Conference, AFIPS Press (Mondale, NJ, 1971), 523-532.

Hansen, W. J., Doring, R., and Whitlock, L. R. Why an examination was slower on-line than on paper. *Int. J. of Man-Machine Studies, 10*, 1978, 507-519.

Hawisher, G. Computers and composition: A critical review. In G. Hawisher and C. Selfe (Eds.), *Coming of Age: Computers in the Composition Classroom.* Teachers' College Press, 1987.

Morris, J., Satyarayanan, M., Conner, M. H., Howard, J. H., Rosenthal, D. S. H., Smith, F. D. Andrew: A distributed Personal Computing Environment. *Comm. ACM*, V. 29, 3 (March, 1986) 184-201.

Muter, P., Latremouille, S. A., Treuniet, W. C., and Beam, P. Extended reading of continuous text on television screens. *Human Factors, 24* (1982), 501-08.

National Science Foundation, *EXPerimental Research in Electronic Submission*. Request for Proposal, 1986.

Rothkopf, E. Z. Incidental memory for location of information in text. *Journal of Verbal Learning and Verbal Behavior 10*, 1971, 608-613.

Shneiderman, B. Direct Manipulation: A Step Beyond Programming Languages, *Computer* V 16, 8 (Aug, 1983), 57-69.

Stallman, R. M. *EMACS: The Extensible, Customizable Self-Documenting Display Editor.* ACM SIGPLAN/SIGOA Symposium on Text Manipulation, 1981.

Wright, P. and Lickorish, A. Proof-reading texts on screen and paper. *Behavior and Information Technology, 2* (1983), 227-235.

Wylie, John, *Personal Editor*, IBM Corporation, 1982.

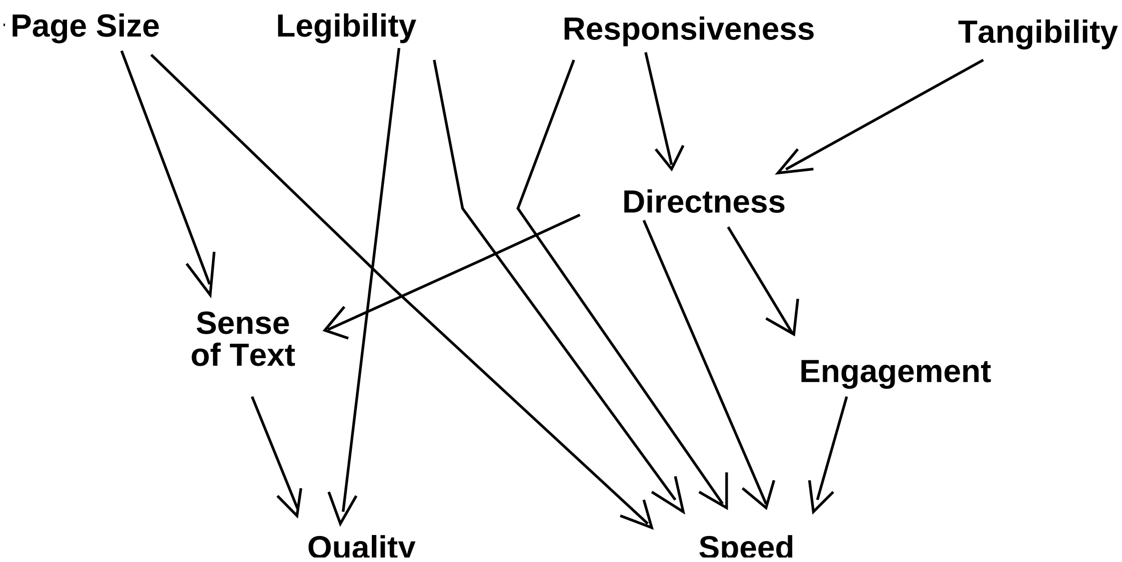

Figure 2. Relationships among the primary and secondary factors. Each line indicates that the upper factor influences the one below. Almost certainly there are many other relationships among these factors.

tab(‘); l l l l l | l ll l l l l | l l l c c c c | c c . _

‘Page‘Legi-‘Respon-‘Tangi-‘Task‘Quality
‘Size‘bility‘siveness‘bility‘Time‘of work

_
paper‘ = ‘ = ‘ = ‘ = ‘ = ‘=

_ **A. Spatial Recall**
PC as terminal‘ = ‘-‘--‘-‘‘- *

_
B. Retrieval
PC as terminal‘--‘-‘--‘-‘-- *‘=
W/S, large window‘ = ‘=-‘-‘=-‘=-‘=

_
C. Reorder Lines
W/S, large window‘ = ‘=-‘-‘=-‘=-‘=
W/S, small window‘--‘=-‘-‘=-‘-- *‘=

_
D. Writing Letters
PC with editor‘--‘-‘-‘-‘ = ‘- *
W/S [1]‘= ‘=-‘-‘=-‘= ‘=

Table 1. Summary of experiments and results. Each computer condition is graded on each of the factors as to whether it is about the same as paper (=), slightly inferior (=-), inferior (-), or very inferior (--). An asterisk (*) indicates a result that is statistically significant at (p<.05) or better.

[1] Results were similar for both large and small windows.

tab(‘); c c s c c s c c c l n n l n n l c s . ‘Mean Time (minutes) ‘_ ‘Scrollbar‘Keys

Large Window‘15.7‘14.4 Small Window‘20.6‘20.7* Paper‘13.5**

Table 2. Mean Time to reorder text. N=10. Method of text advancement was crossed with window size resulting in four computer conditions. (* $p<.01$, ** $p<.05$)

tab(‘); c c c c c c c c l n n n . ‘Time*‘ ‘ ‘(minutes)‘Words*‘WPM

Paper‘13.4‘264‘21 Personal computer‘15.1‘292‘21 Workstation‘17.4‘353‘20

tab(‘); c c s s c l s s c c c c l n n n . ‘Quality ‘_ ‘Content*‘Mechanics‘Total**

Paper‘5.1‘5.7‘10.8 Personal computer‘4.0‘4.3‘8.3 Workstation‘6.0‘5.2‘11.2

Table 3. Results of Letter Writing Experiment. N = 15. Quality was evaluated by graders who sorted responses into four quartiles.

*Differences between highest and lowest number are significant ($p<.05$).

**Differences between lowest number and other two numbers are significant ($p<.05$).

texts for sample screen displays
(Need to be full size font)

Can We Know the Universe?

by Carl Sagan

Science is a way of of thinking much more than it is a body of knowledge. Its goal is to find out how the world works, to seek what regularities there may be, to penetrate to the connections of things--from subnuclear particles, which may be The constituents of all matter, to living organisms, the human social community, and thence to the cosmos as a whole.

Our intuition si by no means an infallible guide. Our perceptions may be distorted by training and prejudice or merely because of the limitations of our sense organs, which, of course, perceive directly but a small fraction of the phenomena of the world. EveN so straightforward a question as whether in the absence of friction of pound of lead falls faster than a gram of fluff was answered incorrectly by Aristotle and almost everyone else before the time of Galileo. Science is based onexperiment, on a willingness to challenge old dogma, on an openness to see the universe as it really is. Accordingly, science sometimes requires courage--at the very least the courage to question the conventional wisdom.

Beyond this the main trick of science is to *really think* of something: teh shape of clouds; the formation of a dewdrop on a leaf; the origin of a man or of a word--Shakespeare, say, or "philantropic"; the reason for human social customs--the incest taboo, for example; how itis that a lens in sunlight can make paper burn; how a "walking stick" got to look so much like a twig; why the Moon seems to follow us as we walk; what prevents us from digging a hole downn to the center of the Earth; what the definition is of "down" on a spherical Earth; or how far is up--does the universe go on forever, or if it does not, is there ayn meaning to the question of what lies on the other side?

Some of these questions are pretty easy. Others, especially the last, are mysteries to which no one even today knows the answer. They are natural questions to ask. Every culture has posed such questions in one way or another. Almost always the proposed answers are in the naature of "Just So Stories," attempted explanations divorced from experiment, or even from careful comparative observations.

But the scientific cast of mind examines the world critically as as if many alternative worlds might exist, as if other things might be here which are not. Then we are forced to ask why what we see is present and not something else. Why are the Sun and the Moon and the pLanets spheres? Why not pyramids or cubes? Why not irregular, jumbly shapes?

If you spend any time spinning hypotheses, checkingto see whether they make sense, whether they conform to what else we know, thinking of tests you can pose to substantiate or deflate your hypotheses, you will find yourself doing science. And as you come to practice this habit of thought more and more you will get better and better at it. To penetrate into into the heart of the thing--even a little thing, a blade of grass, as Walt Whitman said--is to experience a kind of exhilaration that, it may be, only human beings of all the beings on this planet can feel. We are an intelligent species and the use of our intelligence quite properly gives us pleasure. In this rrespect, the brain is like a muscle. When we think well, we feel good. Understanding is a kind of ecstasy.

Hmuan beings are, understandably, highly motivated to find regularities, natural laws. The search for rules, the only possible way to understand such a vast andcomplex universe, is called science. The universe forces those who live in it to understand it. Those creatures who find everyday experience a muddled jumble of events with no predictability, no regularity, are in graVe peril. The universe belongs to those, who, at least to some degree, have figured it out.

It is an an astonishing fact that there *are* laws of nature, rules that summarize conveniently how the world works. And, fortunately for us, we live in a universe that has at least important parts that are knowable. Our common-sense experience and our evolutionary history have prepared us to understand something of thhe workaday world.

How We Listen to Music

by Aaron Copland

We all listen tomusic according to our separate capacities. But, for the sake of analysis, the whole listening process may become clearer if we break it up into its component parts, so to to speak. In a certain sense we all listen to music on three separate planes. For lack of a better terminology, one might name these: (1) the sensuous plane, (2) the expressive plane, (3) the sheerly musical plane. The only advantage to be gained from mechanically splitting up the listening process into these hypothetical planes is the clearer view to be hhad of the way in which we listen.

The simplest way of listening to music is to listen for teh sheer pleasure of the musical sound itself. That is the sensuous plane. It is the plane on which we hear music without thinking, without considering it in any way. One turns on the radio while doing something elseand absentmindedly bathes in the sound. A kind of brainless but attractive state of mind is engendered by the mere Sound appeal of the music.

You may be sitting in in a room reading this book. Imagine one note struck on the piano. Immediately that one note is enough to change the atmosphere of the room--proving that the sound element in musiic is a powerful and mysterious agent, which it would be foolish to deride or belittle.

The suprising thing is that many people who consider themselves qualified music lovers abuse that that plane in listening. They go to concerts in order to lose themselves. They use music as a consolation or an escape. They enter an ideal world where one doesn’t have to think of the realities oF everyday life. Of course they aren’t thinking about the music either. Music allows them to leave it, and they go off to a place to dream, dreaming of and apropos of the music yet nevre quite listening to it.

Yes, the sound appeal of music is a potent and primitive force, but you must not allow it to usurp a disproportionate share of your interest. The sensuous plane is an important one in music, a very important one, but it doesn’t constitute the whole story.

The second plane on which music eXists is what I have called the expressive one. Here, immediately, we tread on controversial ground. Composers have a way of shying away from any discussion of music’s expressive side. Did not Stravinsky himself proclaim that his music was an "object," a "thing," with a life of itsown, and with no other than its own purely musical existence?

This intransigent attitude of Stravinsky’s may be due to the fact that so many people have tried to read different meanings into so mnay pieces. Heaven knows it is difficult enough to say precisely what it is that a piece of music means, to say it definitely, to say it finally so that everyone is satisified with your explanation. But that should not lead one to theother extreme of denying to music the right to be "expressive."

My own belief is that all music has an expressive power, some more and some less, but that all music has a certain meaning behind the notes and that that meaning constitutes, after all, what the piece is saying, what the piece is about. This whole problem can be stated quuite simply by asking, "Is there a meaning to music?" My answer to that would be, "Yes." And "Can yuo state in so many words what the meaning is?" My answer to that would be, "No." Therein lies the difficulty.

Simple-minded souls will never be satisifed with the answer to the second of these questions. They always want music to have a meaning, and and the more concrete it is the better they like it. The more the music reminds them of a train, a storm, a funeral, or any other familiar conception, the more expressive it appears to them. This popular idea of music’s Meaning--stimulated and abetted by the usual run of musical commentator--should be discouraged wherever and whenever it is met. One timid lady once confessed to me that she suspected somethingg seriously lacking in her apprecication of music because of her inability to connect it with anything definite. That is getting the whole thing backward, of course.