Effects of Chunking and Line Length on Reading Efficiency

Stacey A. Keenan

This study investigated whether text with one phrase or "chunk" on each line aids reading in comprehension and proofreading tasks. The chunked format was produced by a computer program that uses syntactic rules to determine chunk boundaries. Contrary to expectations, the chunked format was read significantly more slowly than the standard format in all tasks, at two difficulty levels, and with both screen and paper presentation. A new explanatory variable, line-length variability, could account for these results and also reconcile conflicting findings from previous research. From a literature review, we can infer that both chunking and shorter line lengths have positive effects on reading efficiency, but high line-length variability has a stronger negative effect. Because chunking increases line-length variability, it can interfere with reading.

How can we design text formats to aid reading? This question has generated many hypotheses both from text designers and behavioral scientists. One strategy is to use the findings of psychological research to design text that supports cognitive processes.

One ubiquitous cognitive process is the segmentation and grouping of related information. People segment streams of words into meaningful phrases and sentences as they listen to speech (see Carrithers & Bever, in press, for references) and as they read. Readers pause at the ends of sentences and phrase boundaries (Just & Carpenter, 1980; Mitchell & Green, 1978). These researchers suggest that these pauses correspond to a cognitive process of interclause integration, most likely to link the phrase or sentence to earlier material.

If people need to pause at phrase boundaries for integrative comprehension processing, perhaps reading would be eased if the pause boundaries were visible. One way to reach the goal of designing text to support cognitive processing, then, might be to show phrases typographically. Henceforth, I will call these meaningful phrases "chunks" (Simon, 1974).

Many researchers have investigated whether printing text to show chunk boundaries aids reading. Three methods have been used to show

Visible Language, XVIII 1 (Winter 1984), pp. 61-80.
Author's address: AT&T Bell Laboratories, 190 River Road, Summit, NJ 07901.

chunks: (a) printing only one chunk on each line, (b) printing extra space between chunks (spaced-unit method), but putting as many words and chunks per line as fit, and (c) using two lines for each chunk and printing extra space between chunks horizontally (square-span method).

Several researchers have shown that unskilled readers (children or low-ability adults) comprehend more and read faster with texts that are formatted to have one chunk per line (Cromer, 1970; Grist, 1982; Mason & Kendall, 1979). Other studies have shown that segmented formats improve comprehension and increase reading speed for normal adult readers (Frase & Schwartz, 1979; North & Jenkins, 1951; Dean & Schwartz, 1982). However, Dean & Schwartz found no difference in recall between passages with one phrase per line and passages with short lines (not chunked).

But there is conflicting evidence about whether chunked text aids comprehension for adult readers. Several studies found no advantage for adults reading chunked text (Carver, 1970; Klare, Nichols, & Shuford, 1957). Aaronson & Scarborough (1976), although not testing a chunked format, found that subjects who were reading for comprehension (i.e., to answer a yes-no question) did not pause at phrase boundaries, although subjects who were reading for verbatim recall did. Aaronson and Ferres (1983) concluded that "chunking" is only important when the task is at a syntactic level (e.g., a recall or memorization task), not when the task is mostly at a semantic level (e.g., a comprehension task). Other research has found chunking to help in memory tasks (Anglin & Miller, 1968).

Thus, it seems that chunked formats help unskilled readers and help normal, skilled readers if the task involves memorization or syntactic processing. However, there is some conflict about whether chunked formats can help skilled readers to comprehend more efficiently. Differences in reading materials, tasks, and procedures may account in part for the differences in findings about skilled readers. Frase & Schwartz (1979) used more difficult and complex text than any of the other studies, and their subjects performed a verification task, whereby they read a sentence and then tried to confirm it by reading the text. Because the subjects knew what they were looking for, they might have been able to skip some lines of a chunked format. Other researchers used a multiple-choice task and simpler reading materials.

Other uses for chunked text

People have suggested uses for chunked text aside from reading for comprehension or recall. Some editors say that it is easier to edit text in a chunked format. Many typists type one phrase per line when using word processing equipment, because it makes later search and change operations easier. Because proofreading tasks involve syntactic processing, they probably involve mental segmentation, and hence, chunked formats may be an advantage. As yet, there are no studies testing the effect of a chunked format on proofreading.

Methods of producing chunked formats

With the exception of Grist (1982), all the studies mentioned above relied on human judgment to form meaningful segments or chunks. Boundary judgments are reliable (Frase, Macdonald, & Keenan, in press; Dean & Schwartz, 1982; Johnson, 1970), and Carver (1970) reported that an immediate constituent analysis showed that chunks usually correspond to syntactically significant substrings of sentences. However, some people include "minor breaks" that others exclude (Aaronson & Scarborough, 1976; Klare, et al., 1957). Boundary judgments made by a computer program are consistent (and faster than human judgments). We have such a program at AT&T Bell Laboratories, called *chunk,* which uses syntactic rules to print a text with one phrase per line (Keenan, 1980; Frase, et al., in press). Grist (1982) used this *chunk* program in his study. The segmentation algorithm is based on an analysis of the boundaries marked by a person, and the *chunk* program segments text into chunks that generally agree with those formed by a person. (See the Method section for details.)

Design of the present study

In summary, studies show that at least three factors contribute to reading efficiency: (a) chunking, (b) text difficulty, and (c) task demands. The present study was designed to try to resolve the conflicts among findings of past research about chunked formats, by systematically exploring the relationship among these three factors. To do this, I tested whether segmentation by the *chunk* program aids reading in various tasks with text of two difficulty levels. The tasks ranged from reading for comprehension (semantic processing) to proofreading (syntactic processing).

Because there is evidence that a format with short lines aids reading as much as a chunked format (Dean & Schwartz, 1982), I matched the mean line lengths of the chunked and standard formats. That is, each passage in the standard format had the same mean length as it had in the chunked format. In this way any difference between formats would be attributable to the effect of chunking and not to the effect of mean line length.

If a chunked format does help reading, it would be easy to implement in a screen presentation, but impractical to implement on paper, because it wastes space. This is not a concern for computer-stored text, where any display format is easily and quickly produced from the stored representation. Although video display terminals (VDTs) are becoming widely used, many people still find it unpleasant to read from them. Therefore, I included presentation mode (paper vs. screen) as a factor in this study.

Hypotheses

I hypothesized an interaction between task and format. I expected subjects to perform the proofreading tasks faster when the text was in chunked format than when the text was in standard format. The syntactic editing task, in particular, required that subjects process phrases to get the correct answer. Therefore, I expected subjects to be aided by the chunked format most in that task. In comprehension tasks, readers go beyond the surface or syntactic level, and concentrate on the meaning of the text. Therefore, a syntactically chunked text is less relevant to their purpose. For comprehension tasks, therefore, performance would not differ according to format. However, sentence verification, while involving comprehension, may also involve some syntactic processing. When the reader knows in advance what to look for while reading, he or she may use syntactic cues to choose sections to be read carefully. Therefore, I expected some advantage to the sentence verification task with a chunked format, but not so great an advantage as for the proofreading tasks.

Method

Subjects

Subjects were 24 clerical staff (23 female) from AT&T Bell Laboratories, who volunteered to participate. All had completed high school; none had had more than two years of college. Their mean age was 34.

Reading Materials

Each text passage was about one paragraph long and had about 130 words. The easy passages were from *Adventures in Living Plants,* a sixth grade text by Edwin B. Kurtz, Jr., and Chris Allen, published in 1965 by the University of Arizona Press. Portions of the book were reproduced with permission from the publisher. Although the paragraphs from this book were easy to read, they comprised detailed facts on botany. The reading grade level for all easy paragraphs combined was 4.8, as measured by the Kincaid readability formula (Coke, 1978). Coke (unpublished paper, 1978) showed that although the Kincaid formula tends to underestimate the reading grade level of easy text, it is more accurate than the Flesch formula, the Automated Reading Index, and the Coleman-Liau formula are on easy text. She also found the Kincaid formula to be the most accurate for measuring difficult text.

The difficult passages were from *The Machinery of the Brain,* a college level book by Dean E. Wooldridge, published in 1963 by McGraw-Hill. Portions of the book were reproduced with permission of the publisher. The paragraphs from this book described the physiology of the brain and some animal behavior that is controlled by the brain. The reading grade level for these paragraphs combined was 16.8, as measured by the Kincaid readability formula. Both books presented much detailed information that was not likely to be known by the subjects. Indeed, after the experiment, subjects reported no special training in or knowledge of the material presented.

The practice passages were selected randomly from a set of reading materials, of varied subject matter and medium difficulty, which have been used in other reading studies. Most of these paragraphs were factual and all were suitable for questioning. The Kincaid reading grade level of these passages taken together was 9.9.

Presentation

The passages were presented in equivalent formats either on a Hewlett-Packard 2645A VDT (dark background, light blue letters) or printed with a daisy wheel printer on white paper. Both displays had the same number of lines on each page or screen, and the same number of words on each line.

The chunked format was produced by the *chunk* program. The standard format was produced by a text formatting program with the line length for each passage set to be the same as the average length of the lines in the chunked format for that passage. However, the right margin was unjustified so that not every line was the same length. Examples of the chunked and standard formats are shown in Figures 1 and 2.

Chunking algorithm

Development of the algorithm for the *chunk* program was based on an analysis of how people mark phrases in text. The program breaks text after certain punctuation marks, and before or after certain words which have been identified as break words. Most break words are prepositions and conjunctions. These breakpoints are arranged in levels so that only the highest-level or strongest will cause a break when two or more breakpoints are close together.

The *chunk* program was written several years ago at AT&T Bell Laboratories. Its performance was tested in a small experiment with five human judges. There was 90% agreement among the five judges that the chunk boundaries marked by a person demarcated meaningful units of information. In addition, we found high agreement between the chunk boundaries marked by that person and those marked by the *chunk* program. The *chunk* program chose 73% of the same chunk boundaries that a person chose for 60 passages of easy, medium, and difficult text. Another measurement method, used by Johnson (1970), considers each word to be a decision point (i.e., Does

Figure 1. Example of the chunked format.

Things happened so fast I forgot
to tell you about the special chemicals
that make the cell wall stretch like bubble-gum.
The chemicals are called auxins.
Auxins are made by cells
that are rapidly dividing.
Then the auxins move down the stem
to where cells are ready to enlarge.
And then they cause the cells
to grow longer.
Auxins are very potent chemicals in plants.
A small ball of auxins the size
of a pinhead
is enough chemical to start millions
and billions of cells grower.
Three men discovered auxins —
two chemists and a botanist.
The botanist, Dr. Fritz Went,
discovered a way
of measuring the effect of auxins
on plants.
The chemists, Dr. Kogl and Dr. Haagen-Smit,
identified the first auxin.

END

this word end the chunk?). Calculated in this way, the *chunk* program made 89% of the same word-by-word boundary decisions as a person did for the 60 test passages. A more balanced measure, the geometric mean of the word-by-word agreement proportions, shows 83% agreement between the *chunk* program and a person's judgments.

Tasks

Four reading tasks were chosen to represent a variety of reading skills, from high-level comprehension to lower-level proofreading. The tasks were: recognition, sentence verification, syntactic editing, and search for misspellings. The first two tasks can be classified as comprehension tasks, as the subject was required to answer a question based on what she or he read in a passage. The latter two tasks can be classified as

Figure 2. Example of the standard format.

Things happened so fast I forgot to
tell you about the special
chemicals that make the cell wall
stretch like bubble-gum. The
chemicals are called auxins.
Auxins are made by cells that are
rapidly dividing. Then the auxins
move down the stem to where cells
are ready to enlarge. And then
they cause the cells to grow
longer. Auxins are very potent
chemicals in plants. A small ball
of auxins the size of a pinhead is
enough chemical to start millions
and billions of cells grower.
Three men discovered auxins — two
chemists and a botanist. The
botanist, Dr. Fritz Went,
discovered a way of measuring the
effect of auxins on plants. The
chemists, Dr. Kogl and Dr. Haagen-
Smit, identified the first auxin.

END

proofreading tasks, as they required the subject to find a single word error in each passage. These tasks are described below.

1 *Recognition.* Subjects read a passage, and then answered a multiple-choice question about the passage. Subjects had to understand at least the gist of the passage in order to answer correctly. Simply recognizing that phrases occurred in the passage was not enough, because phrases from the passages were used in both correct and incorrect choices.

2 *Sentence verification.* Subjects read a sentence, which was either true or false; then they read a passage to verify the sentence. Verification required information from several sentences to be integrated. Subjects were instructed to read the entire passage and then to re-read the true-false sentence, before giving an answer, even if they thought they knew the correct answer. In this way, each subject read the same number of words and reading rate could be measured. This task was simpler than the recognition task, because subjects knew what information they were looking for as they read. They did not need to understand the entire passage in as great depth as they did in the recognition task. Half of the sentences at each difficulty level in this task were true, half were false.

3 *Syntactic editing.* Subjects read a passage looking for syntactic errors. The errors were single words with the ending changed to make the word the wrong part of speech, the wrong tense, or the wrong number (i.e., plural vs. singular). However, the incorrect word was still a real word. For example, in the following sentence the word "formation" should be the word "formed."

> They are formation near the tip of the root.

The passage in Figures 1 and 2 also contains a syntactic error; the word "grower" should be the word "growing." Four of every five passages contained such a word; one passage at each difficulty level had no error. Since these errors were easy to miss (for example, one might read the above example as "They are *a* formation near the tip of the root"), passages without errors were included to discourage subjects from reading passages twice if they did not see the error on the first reading. Hence, subjects did not know if they answered incorrectly. This made the task equivalent to the recognition and sentence verification tasks in that respect.

The subjects were informed that about one out of every five passages would not have a mistake. They were told to read carefully and not to read any passage twice. Only the four scores for passages with errors were included in the data analysis. Thus there were four scores for each difficulty level in each of the four tasks.

Although this is a proofreading task, subjects could not find the error by reading words in isolation. Since each error was still a real word, the errors could only be found when read in the sentence or phrase context. So it was a lower-level task than either of the comprehension tasks, in that it required little comprehension. However, it was a higher-level task than the next one, search for misspellings, because it required syntactic processing of whole sentences or phrases.

4 *Search for misspellings.* Subjects were asked to read a passage, looking for misspelled words. As in the syntactic editing task, four of every five passages contained an error (misspelling) and one passage at each difficulty level had no error. The subjects were informed that about one out of every five passages would not have a mistake. Only the four scores for passages with errors were included in the data analysis.

In all tasks, subjects were given instructions for the task they were to perform before they began reading. Subjects were given four practice trials for each task type, except for the syntactic editing task. A pilot test showed that subjects needed more practice with the syntactic editing task, so there were eight practice trials for that task.

Design

A 2 × 4 × 2 × 2 (Format × Task × Text Difficulty × Presentation Mode) factorial design was used, with repeated measures on task and text difficulty. In other words, each subject performed all four tasks, at both difficulty levels. Format and presentation mode were between-subjects measures. Subjects were randomly assigned to one of four conditions: paper-chunked, paper-standard, screen-chunked, and screen-standard, with six subjects in each group. Subjects read eight passages for each of the two comprehension tasks, and 10 passages for each of the two proofreading tasks. The sequences of tasks and passages were randomized. Within each task, half of the passages were easy, and half were difficult.

Scoring

In tasks such as the ones used in this experiment, subjects can usually sacrifice accuracy for a faster performance or vice versa (Wickelgren, 1977). To avoid the possible confounding of a speed-accuracy trade-off, reading speed was the only dependent measure, while the accuracy was held relatively constant. Subjects were instructed to read carefully because it was more important for them to get the correct answer than to read very quickly.

Reading rate for the recognition and sentence verification tasks was calculated by dividing the time spent reading a passage by the number of characters (including spaces) in the passage. For the two proofreading tasks, reading rate was calculated by dividing the reading time by the number of characters preceding the answer. I recorded reading times with a digital stopwatch.

Answers were coded in one of three categories: right, wrong, or "wrong, but." The third category was created for answers which were wrong, but did not indicate that subjects were not reading carefully. For five questions (out of 32) the same wrong answer was given by nearly all subjects. This indicates that these questions were too hard for the subjects or were poorly worded, but not that the subjects were not reading carefully. The wrong answer given by most subjects to these questions received the "wrong, but" categorization. In addition, in the syntactic editing and search for misspellings tasks, occasionally subjects would find a word preceding the target that they thought was used or spelled incorrectly. This type of mistake was also scored as "wrong, but." Right answers were scored as 1.0, wrong answers were scored as 0.0, and "wrong, but" answers were scored as 0.5. Thus, the means for errors can be interpreted as the proportion of answers which were correct.

Results

Accuracy. Although accuracy was not the dependent measure, it was analyzed to see if subjects were trading reading speed for accuracy, or vice versa. The mean proportion of correct answers was .81, meaning that 81% of all the answers given were correct. A $2 \times 2 \times 4 \times 2$ (Mode $\times$ Format $\times$ Task $\times$ Difficulty) split-plot analysis of variance was conducted, with subjects nested in mode-format groups and subjects crossed with task and difficulty. Subjects were treated as a random variable. Two main effects were significant: task, $F(3,60) = 2.90, p < .05$, and difficulty, $F(1,20) = 20.86$, $p < .01$. The main proportion of correct answers was .86 for the easy text and .76 for the difficult text. Subjects performed most accurately on the search for misspellings and recognition tasks (proportion correct = .86 and .83, respectively), and somewhat less accurately on the syntactic editing and sentence verification tasks (proportion correct = .77 and .78, respectively). Because subjects in different mode-format groups performed with equal accuracy, we may use reading rate as the dependent measure in the rest of the analysis. Subjects in different mode and format groups did not trade accuracy for speed in different ways.

Reading Rate. All subjects varied greatly in reading rate, but one subject in the screen-chunked group read much faster than any other subject (up to 50 characters per second for some passages) and answered only half of the questions correctly. Her scores were eliminated and replaced by the means of the other five subjects in the screen-chunked group.

The same split-plot analysis of variance design described above for accuracy analysis was used to analyze reading rate. Reading rate was measured in characters per second. The overall mean was 17.32 characters per second (and 2.97 words per second, or 178 words per minute). Reading times, like reaction times, are often positively skewed when the measured times are short. In this case for example, no subject read less than 5 characters per second. Therefore, there was not much variation in reading rates for subjects who were slower than average, but faster subjects ranged up to 40 characters a second for some passages. I performed a logarithmic transformation on the data, as is recommended with skewed measures (Myers, 1966).

Contrary to expectations, the chunked format was read at a significantly slower rate than the standard format, $F(1,20) = 5.78$, $p < .05$). Subjects read the chunked format at a rate of 15.8 characters per second and they read the standard format at a rate of 18.85 characters per second. In other words, the chunked format caused a 16% reduction in reading speed. Mode had no significant effect on reading rate ($F(1,20) < 1.0$), and no interactions involving mode or format were significant.

The main effects of difficulty and task were significant, as expected. The easy text was read faster than the difficult text, $F(1,20) = 21.83$, $p < .01$. Subjects read the easy text at a rate of 17.92 characters per second and the difficult text at a rate of 16.72 characters per second. Task also had a significant effect on reading rate, $F(3,60) = 57.26$, $p < .01$. The comprehension tasks were read faster than the proofreading tasks. The reading rates for each task in characters per second (cps) were: sentence verification = 20.5 cps; recognition = 19.1 cps; search for misspellings = 16.0 cps; and syntactic editing = 13.7 cps.

Tests of the strength of association for the significant effects showed that format accounts for 8% of the variance, task accounts for 32% of the variance, and difficulty accounts for 2% of the variance. Thus, the significant effects account for a total of 42% of the variance in reading rates. The great deal of variability between subjects probably accounts for most of the rest.

Discussion

The results are quite different from those I hypothesized. Not only did the chunked format cause slower reading, but task did not interact with format. That is, the chunked format was detrimental to the performance of all tasks, whereas I had expected the chunked format to ease significantly performance of only some tasks. Even past studies that found no beneficial effect of chunking did not find chunking to be so detrimental to reading efficiency.

To find an explanation for these contradictory results, we must look at how this study differs from past studies. The formats used in this study differed from the formats used in others in two important ways. First, I used a chunked format that was produced by a computer program, rather than by human judges. Second, I matched the average line lengths of passages in both formats, so that the standard format had shorter lines than the standard used in other studies. I will now consider the possibility that computer-chunked text or matched line lengths produced these surprising results.

Were the chunks unacceptable?

One possibility is that although chunking may help readers, *chunk's* chunks may not be good enough. Frase & Schwartz (1979) showed that a misapplied segmentation strategy caused slower reading than standard text. If the chunks produced by the *chunk* program represent a poor segmentation strategy, then one would expect the chunked format to cause slower reading. Although the chunked format did produce slower reading, there is some evidence that *chunk's* chunks are like those people produce (see the section on the chunking algorithm in the Method section). However, without a more stringent test of *chunk's* acceptability, we cannot conclude whether the *chunk* program's chunks were so poor as to produce the slow reading. However, it seems unlikely that this is the case.

Was the chunked format too variable in line length?

Let us now consider the possibility that the line lengths, and not the chunks, caused the unexpected results. When one looks at the two formats (see Figures 1 and 2), the striking difference between them is that the chunked format has such a jagged right margin. Although the mean line length was held constant, the line-length variability necessarily was not. The overall mean line length for the chunked text was 30 characters per line, and was 32 characters per line for the standard text. However, the standard deviations of the line lengths of the standard passages averaged 2.6 characters, whereas the standard deviations of the line lengths of the chunked passages averaged 10.2 characters.

Effect of line-length variability on reading

Extreme variability in line length may slow reading by disrupting the rhythm of eye movements, in particular, of return sweeps. A return sweep is the movement of the eye from the end of one line to the beginning of the next. They are usually somewhat inaccurate; "undershoots" often occur, which require the eye to make a corrective regressive movement to get to the beginning of the line (O'Regan & Levy-Schoen, 1979). Return sweeps take from twice as long to five times as long to make as ordinary saccadic movements. Leisman (1978) hypothesizes that when lines of text are both right- and left-justified, the reader builds an internal map with the coordinates of the beginning and ending of lines. Then the brain constructs a "program" to execute the return sweep automatically. If this explanation is true, the extreme unpredicability of the distance from the right margin to the left margin in the chunked format would make such an automatic program impossible for a reader to set up. Even if Leisman's hypothesis about automatic return sweeps is not true, variable lines still might cause return sweeps to be inaccurate more often. If so, reading speed must decrease because corrective saccades take as long to make as normal saccades do.

Effect of mean line length on reading

The present study differed from most previous studies (except Carver, 1970) by holding mean line length constant across formats. Subjects in both format groups read text with medium-length lines (about 3 inches wide, averaging 5 words per line). Tinker (1963) cites evidence that people prefer to read lines of about 14-31 picas (2-5 inches) in width and that they also read such widths more quickly. (Tinker's width recommendations vary with point size and leading. These are for standard 10-point type with 2-point leading.) Because most previous studies did not control for line length, we cannot tell if, when subjects read more efficiently with a chunked format, they were responding to the presence of visible chunks, or to the shorter lines.

Besides being preferred by readers, medium-length lines may help reading eye movements. For example, return sweeps may be more accurate when lines are of medium length, because the eye can see the beginning of the next line in peripheral vision, when it is at the end of a line. Tinker (1965) reports that regressive eye movements increased by over 50% when text had long lines (43 picas). He attributes this increase to readers making inaccurate return sweeps, and the subsequent need to make a corrective regressive movement. A revised theory of how to improve text design needs to consider eye movement limitations in addition to aids to cognitive processing. That is, a potential aid to comprehension, such as placing one chunk per line, should not be allowed to disrupt eye movements, or the cognitive advantage may be offset by the disruption.

Hence, the unexpected results of this study suggest an explanation for previously conflicting experimental results. Below, I consider three critical variables in classifying text formats: chunking, mean line length, and line-length variability. These three together might allow us to resolve contradictory findings of other research. Presumably, a chunked format aids reading, and a moderate line length with little variability aids eye movements.

Reinterpretation of studies with chunked formats

Most studies, including this one, have ignored a potentially important characteristic of text format — that of line-length variability. Most past studies with chunked formats have also ignored line length as a variable. Another look at the results of past studies shows that if these variables are considered, most conflicts might be resolved. Table I shows a reinterpretation of the results of this and three previous studies on the effect of chunked formats on the reading efficiency of skilled readers. The new interpretation considers three format variables, chunking, line length, and line-length variability — whether the experimenters specifically included all three variables or not. Chunking and medium mean line length are considered to be advantages; high variability in line lengths is considered a disadvantage to efficient reading. The first column of Table I names the study and the second column names the formats that are being compared (in the original experimenter's terms). The third column names which variables differ between the formats being compared, whether they were explicitly studied or not. The fourth column shows which format was read faster. The fifth column shows what we can conclude from the results of the comparison in light of a three-variable theory.

Frase & Schwartz (1979) compared five formats. Their meaningfully-segmented, meaningfully-indented format (MI) is, in my terminology, a chunked format. They found indentation to be only a weak aid to reading; therefore, I will include it as a variable in Table I, but I will not dwell on it. Their other three non-standard formats were permutations of chunking and indenting with misapplied segmenting and indenting strategies. All four of these experimental formats had

Table I. Reinterpretation of Chunked Format Studies Using Three-variable Theory

Study	Comparison	Variables which differ	Result	Conclusion
Frase & Schwartz (1979)	S vs. MI	chunking, line length, ll variability, indentation	MI 18% faster	chunk + indent + med. lines > variab.
	S vs. MNI	chunking, ll variability, line length	MNI 12% faster	chunk + med. lines > variab.
	S vs. NMI	line length, ll variability, indentation	S = NMI	med. lines + indent = variab.
	S vs. NMNI	line length, ll variability	S 16% faster	med. lines < variab.
	(MI + MNI) vs. (NMI + NMNI)	chunking	M's > NM's	chunking = advantage
North & Jenkins (1951)	S vs. spaced-unit	chunking	spaced-unit faster	chunking = advantage
Carver (1970)	chunking vs. "newspaper"	chunking, ll variability in-line variab.	chunk = newspaper	chunk – in-line variab. = variab.
Keenan (1984)	S vs. chunking	chunking,	S 19% faster	chunk < variab.

S = standard format
MI = meaningfully-segmented, indented format
MNI = meaningfully-segmented, nonindented format
NMI = nonmeaningfully-segmented, indented format
NMNI = nonmeaningfully-segmented, nonindented format
ll variability = line length variability

medium lines that varied in length as compared to the standard format (S), which had a right margin at 66 characters. The first row in Table I shows the comparison of the chunked format (MI) and the standard format (S). The MI format was chunked, had medium-length lines, was meaningfully indented, and had high line-length variability. The format characteristics of the S format were opposite to those of the MI format. Frase & Schwartz found that subjects read text in the MI format 18% faster than they read text in the S format. From this we can conclude that the combination of the advantages to reading of chunking, indentation, and medium-length lines outweigh the disadvantage of extremely variable line lengths. Because chunking, indentation, and line length are confounded, we cannot conclude simply that chunking aids reading. The second row in Table I shows that the combination of the advantages of chunking and medium-length lines (without the advantage of indentation) are enough to outweigh the disadvantage of variable line lengths. In the third row we see that without the advantage of chunking, medium-length lines and indentation do not outweigh line-length variability, but equal it.

The fourth row of Table I shows the comparison of the improperly segmented, medium-length line format (NMNI) and the standard format (S). These two formats differ only in length (the NMNI format had shorter lines) and in line-length variability (the NMNI format had variable line length). Because subjects read the standard format 16% faster, we can conclude that the single advantage of having shorter lines is outweighed by the disadvantage of having variable line lengths.

Frase & Schwartz did one comparison which did not confound line length and line-length variability with chunking. The fifth row in Table I shows the comparison of the two chunked formats with the two non-meaningfully segmented formats. All of these formats had variable, medium-length lines. Because the chunked formats were read significantly faster than the others, we conclude that a chunked format does aid reading, when all other things are equal.

The spaced-unit format used by North & Jenkins (1951) did not introduce line-length variability or shorter lines, because it did not place one chunk on each line. Instead, the spaced-unit format resembles the standard format in every way, except that there are extra spaces placed at chunk boundaries. Hence, their study really did show that chunking is an advantage in reading for comprehension.

Carver (1970) found no difference in reading speed or comprehension between a chunked format and a "newspaper" format. Carver's newspaper format was similar to the standard format in the present study, in that the lines were as short as chunked lines, but his format was right-justified. Although a right-justified format eliminates line-length variability, it also introduces within-line variability, because extra spaces are inserted between words to force the last word on the line to end at the right margin. Campbell, Marchetti, & Mewhort (1981) show that unpredictable spaces between words (produced by a fixed-space, right-justification technique like the one Carver used) reduce reading speed about as much as unpredictable right margins do. From Carver's study we may conclude that the advantage of a chunked format with uniform in-line spacing over a non-chunked format with within-line variability, is offset by the disadvantage of variable line lengths in the chunked format.

As shown in the last line of Table I, the present study confounded the effects of chunking with line-length variability. That subjects read the standard format 19% faster than the chunked format shows that the advantage of chunking is outweighed by the disadvantage of line-length variability.

In summary, this analysis presents evidence for three findings:

1 Chunking appears to be an advantage when line length and variability are held constant.

2 Medium-length lines are an advantage in reading when chunking and line-length variability are held constant.

3 Low line-length variability (i.e., a regular margin) may be an advantage in reading.

In addition, line-length variability seems to have the strongest influence on reading efficiency. As Table I shows, neither chunking nor shorter lines alone provides an advantage great enough to override the disadvantage of having variable line lengths. Eye movement disruptions seem to be more detrimental to reading speed than chunking is helpful to comprehension.

Practical implications

If eye movement disruptions are the reason for the chunk disadvantage, then chunked formats, as they have been implemented to date, do not provide any advantage to reading over shorter line standard formats, such as formats found in newspapers. However, future chunk algorithms might try for less variability. A format between chunked and medium-length line standard might be better than either. Such an algorithm would always start a sentence on a new line, as *chunk* does now. In addition, it would need to consider whole sentences at a time in deciding where to break lines. Most sentences, especially when the text is difficult, have more possible chunk boundaries than are used. A smarter program would look ahead to see which breaks would result in the least variation in length. Plass & Knuth (1982) describe a line-breaking algorithm that considers a whole paragraph at a time rather than a single line at a time in choosing breaks, and minimizes "badness" of breaks across the paragraph, by assigning penalties to breaks that are considered undesirable for one reason or another. A new chunking algorithm could work similarly, assigning a penalty of zero to all possible chunk boundaries and assigning high penalties to all other possible boundaries. Minor chunk boundaries might be assigned low penalties, rather than zero penalties. The highest penalties would go to spaces between words which should not be split up. Demerits would be assigned to lines that are outside variability limits. The new *chunk* program would then create chunks to minimize penalties, and also to minimize line-length variability.

Conclusion

The unexpected results of this study clarify the contribution of line-length variability to reading efficiency. When variability is considered, the conflicts among results from studies of chunked formats are resolved. A chunked format does ease reading, and so does a format with medium-length lines. But these advantages are not so strong as the advantage of little variation in line length. High variability in line lengths disrupts eye movements enough to outweigh the benefits of a chunked format with medium-length lines.

Further studies are needed to confirm these conclusions. A study similar to the present study, but which includes line-length variability as an independent variable, would show the individual and combined effects of chunking, mean line length, and line-length variability on reading. Studies of eye movements when the reader is reading text with lines of varying lengths would show whether inaccurate return sweeps are the cause of slower reading in such formats. Finally, further studies of the acceptability of the chunks produced by the *chunk* program (and a new *chunk* program, which minimizes variability), are needed before entirely acceptable chunked formats can be produced.

References

Aaronson, D., & Ferres, S. (1983). A model for coding lexical categories during reading. *Journal of Experimental Psychology: Human Perception and Performance, 9,* 700-725.

Aaronson, D., & Scarborough, H. (1976). Performance theories for sentence coding: Some quantitative evidence. *Journal of Experimental Psychology: Human Perception and Performance, 2,* 56-70.

Anglin, J.M., & Miller, G.A. (1968). The role of phrase structure in recall of meaningful verbal material. *Psychonomic Science, 10,* 343-344.

Campbell, A.J., Marchetti, F.M., & Mewhort, D.J.K. (1981). Reading speed and text production: A note on right-justification techniques. *Ergonomics, 24,* 633-640.

Carrithers, C., & Bever, T.G. (in press). Eye-fixation patterns during reading confirm theories of language comprehension. *Cognitive Science.*

Carver, R.P. (1970). Effect of a chunked typography on reading rate and comprehension. *Journal of Applied Psychology, 54,* 288-296.

Coleman, E.B., & Hahn, S.C. (1966). Failure to improve readability with a vertical typography. *Journal of Applied Psychology, 50,* 434-436.

Cromer, W. (1970). The difference model: A new explanation for some reading difficulties. *Journal of Educational Psychology, 61,* 471-483.

Dean, R.S., & Schwartz, N.H. (1982). Inducing comprehension in the inconsistently lateralized reader. Manuscript submitted for publication.

Frase, L.T., Macdonald, N.H., & Keenan, S.A. (in press). Intuitions, algorithms, and a science of text design. In T. Duffy & R. Waller (Eds.), *Designing usable text.* New York: Academic Press.

Frase, L.T., & Schwartz, B.J. (1979). Typographical cues that facilitate comprehension. *Journal of Educational Psychology, 71,* 197-206.

Grist, S. (1982). *Effect of discourse segmentation on comprehension and reading rate of eighth-grade children.* Unpublished doctoral dissertation, Rutgers University, New Brunswick, NJ.

Johnson, R.E. (1970). Recall of prose as a function of the structural importance of the linguistic units. *Journal of Verbal Learning and Verbal Behavior,* 12-20.

Just, M.A., & Carpenter, P.A. (1980). A theory of reading: From eye fixations to comprehension. *Psychological Review, 87,* 329-354.

Keenan, S.A. (1980, April). Computer projections of the cognitive effects of text changes. In L.T. Frase (Chair) *Computer aids for writing and text design.* Symposium conducted at the Annual Meeting of the American Educational Research Association, Boston.

Kirk, R.E. (1982). *Experimental design: procedures for the behavioral sciences* (second edition). Monterey, CA: Brooks/Cole Publishing Company.

Klare, G.R., Nichols, W.H., & Shuford, E.H. (1957). The relationship of typographic arrangement to the learning of technical training material. *Journal of Applied Psychology, 41,* 41-45.

Kurtz, E.B., & Allen, C. (1965). *Adventures in living plants.* Tucson: University of Arizona Press.

Leisman, G. (1978). Ocular-motor system control of position anticipation and expectations: Implications for the reading process. In J.W. Senders, D.F. Fisher, & R.A. Monty (Eds.), *Eye movements and the higher psychological functions,* (pp. 195-207). Hillsdale, NJ: Lawrence Erlbaum Assoc.

Mason, J.M., & Kendall, J.R. (1979). Facilitating reading comprehension through text structure manipulation. *The Alberta Journal of Educational Research, 25,* 68-76.

Mitchell, D.C., & Green, D.W. (1978). The effects of context and content on immediate processing in reading. *Quarterly Journal of Experimental Psychology, 30,* 609-636.

Myers, J. L. (1966). *Fundamentals of Experimental Design.* Boston: Allyn and Bacon.

North, A.J., & Jenkins, L.B. (1951). Reading speed and comprehension as a function of typography. *Journal of Applied Psychology, 35,* 225-228.

O'Regan, K., & Levy-Schoen, A. (1979). The control of eye movements in reading (tutorial paper). In P.A. Kolers, M.E. Wrolstad, & H. Bouma (Eds.), *Processing of visible language* (Vol. 1, pp. 7-36). New York: Plenum Press.

Plass, M.F., & Knuth, D.E. (1982). Choosing better line breaks. In J. Nievergelt, G. Coray, J.D. Nicoud, & A.C. Shaw (Eds.), *Document preparation systems* (pp. 221-242). New York: North-Holland Publishing Company.

Simon, H.A. (1974). How big is a chunk? *Science, 183,* 482-488.

Tinker, M.A. (1963). *Legibility of print.* Ames: Iowa State University Press.

Tinker, M.A. (1965). *Bases for effective reading.* Minneapolis: University of Minnesota Press.

Wickelgren, W.A. (1977). Speed-accuracy tradeoff and information processing dynamics. *Acta Psychologica, 41,* 67-85.

Wooldridge, D.E. (1963). *The machinery of the brain.* New York: McGraw Hill.

Author Notes. This paper is based on a master's thesis prepared for New York University, "Effects of Display on Reading Tasks." Special thanks go to my colleague and thesis advisor, Larry Frase, who contributed much advice, discussion, and encouragement.