Drawing from the conclusions of research studies by linguists and psychologists, this paper describes four classes of principles for constructing readable prose for both print materials and computer screens: (1) global principles dealing with a schema-theory approach to discourse type, with topical focus, and with parallelism; (2) sentence-level principles pertaining to verbal style, noun strings, conditional statements, right branching, negatives, active voice, and pronoun deletion; (3) word-level principles pertaining to length, frequency, and concreteness; and (4) graphics principles dealing with space, lines, and highlighting. Examples are offered of how these principles might be applied to the design of instructional materials or to teaching methods. Fifty-five references are listed. (Author/LMM)
Designing Text for Information Processing

TR 86  June 17, 1984

TR 011880

Patricia A. Milazzo

PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY Patricia A. Milazzo TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
DESIGNING TEXT FOR INFORMATION PROCESSING

Ann Humes

ABSTRACT

Three classes of principles are described for constructing readable prose for both print materials and computer "screens": global principles dealing with a schema-theory approach to discourse type, with topical focus, and with parallelism; sentence-level principles pertaining to verbal style, noun strings, conditional statements, right branching, negatives, active voice, and pronoun deletion; word-level principles pertaining to length, frequency, and concreteness; graphics principles dealing with space, lines, and highlighting.
DESIGNING TEXT FOR INFORMATION PROCESSING

Ann Humes

Beyond fourth grade, students read primarily to learn about subject areas such as science, history, or health. The readability of prose that provides information students must process when they are "reading to learn" is an important concern in designing text that extends over several paragraphs or pages. The designer of this text wants to write readable prose, but the task requires more than just using well-known words and short sentences. Although vocabulary and syntax should be controlled so that students are not overwhelmed by what may be unfamiliar, the designer must be allowed a margin of difficulty in the text so that it transfers information completely and explains new ideas and concepts adequately.

Over the years, a sizable body of research has been conducted to define variables that affect the readability of prose intended for young learners whose reading skills are still maturing. This research has taken into account not only such variables as vocabulary, sentence length, and text structure, but also such features as the arrangement of text on the printed page. Recently, with the introduction of computers as devices for presenting text, considerable new research has explored ways to improve the readability of prose that is displayed on computer screens.

Readable prose is defined here as text with meaning that can be "easily and quickly comprehended for an intended purpose by an intended reader operating under normal conditions of alertness, motivation, time pressure, etc." (Huckin, 1983, p. 103). Furthermore a readable text is
one that "not only conveys its meaning efficiently but also facilitates recall of that meaning" (p. 102).

Developers of instruction most commonly have measured readability by using formulas that count the number of syllables, words, and sentences in a text. Popular formulas appropriate for elementary school materials include the Spache formula (for students fourth grade or below), the Dale-Chall formula (for students fifth grade and above), the Fry formula, the Raygor formula, and the Fog and Smog formulas. Since there is some evidence that short sentences make texts more readable (e.g., Coleman, 1982), these tests should be used to screen text and thus ensure that the "readability level" according to standard formulas is not excessively high.

However, these formulas have shortcomings. Studies indicate that the formulas predict readability, as defined by tested linguistic criteria, in only 50% of the cases (Davison, Lutz, Roalef, 1981). Various studies have demonstrated that other features are more important than sentence length (e.g., Kintsch and Keenan, 1973; Holland, 1981). Furthermore, studies have also shown that shorter sentences may make a text more difficult to comprehend because the text presents related ideas separately, forcing the reader to infer the semantic connections (e.g., Davison et al, 1981; Fodor, Bever, & Garrett, 1974). The following examples demonstrate the problem; the sentences in example A do not provide linguistic connections, but those in B do.

A. I like to telephone.  
   I want to telephone Sam.  
   I dial his number.  
   Sam's telephone rings.  
   I hear a noise in my telephone.  
   Sam's telephone is ringing.
B. I like to telephone. I want to telephone Sam, so I dial his number. Then a noise in my telephone tells me that Sam's telephone is ringing.

Readability formulas would rate A as more readable, yet the absence of linguistic links between ideas that are present in B make B more comprehensible (Davison, et al, 1981).

Features that linguists and psychologists have identified as important for readability should be considered in the design of text for students, whether that text is for print materials or for computer screens. Consequently, these principles are discussed below. The explanations, with examples, are brief because a complete treatment would include both a detailed discussion of all the research and more topics than are necessary to consider here.

Since these principles should be used as guidelines for composing appropriate text, they are presented in the approximate order of choices that a writer must make. A writer first makes global choices of schema before he or she composes sentences and selects individual words. Following these principles is a discussion of appropriate screen displays in so far as screen design is a factor in readability.

The principles discussed below are matters of technology rather than art—they have been explored and supported by research studies. Some, but not all, of these studies are cited in the discussion. Furthermore, global concerns are treated more extensively because they are more complex, yet less familiar than are the sentence, word, and screen concerns.
GLOBAL PRINCIPLES FOR WHOLE TEXTS AND PARAGRAPHS

An individual's long-term memory stores patterns of information called schema (Huckin, 1987). Schema appear to contain both perceptual information and relational information, as well as procedural information; they are stereotyped scripts that are activated during the comprehension process (Freedle & Fine, 1982). Exposure to different kinds of texts develops in students' minds the schema for different discourse types. Then, as a person reads, cues in the text to the appropriate schema summon that schema from the reader's long-term memory.

This phenomenon has two contrasting effects on the readability of instructional text: students who do not have a schema for the type of discourse employed in instruction will have more difficulty both comprehending and recalling information in the discourse (Freedle and Fine, 1982); however, if students do have the schema stored in memory, the salient features of the schema will help their comprehension and recall (e.g., Mandler and Johnson, 1977). In fact, Taylor and Samuels (1983) found that students who had superior recall were using discourse schema rather than memory to enhance that recall.

The first discourse-type schema that children store in their long-term memory is the schema for narrative discourse, for sequential activities or events. Britton, Graesser, Glynn, Hamilton, and Penland (1983) found in a series of six experiments that passages cast in narrative discourse produced more meaning than those comprised of expository discourse because "schemata for narrative texts are more familiar, more frequently encountered, and easier to comprehend than
schemata for expository texts" (p. 39). Furthermore, narrative text is so much easier that some readers "rewrite" non-narrative discourse as they read: Flower, Hayes, and Swarts (1980) discovered that many students in their study created narratives to explain text to themselves and thus remember definitions, principles, and concepts. Other studies have shown that when documents were revised from expository to narrative prose, they were more readable (e.g., Gunnarson, 1981).

Because this schema transformation can increase the readability of a text, it should be considered in designs whenever it is feasible. For example, Freedle and Fine (1982) suggest that science texts sometimes include long passages that describe what goes into producing a chemical reaction and that these passages can be shaped into a saliently narrative form. They also suggest that changing present tense verbs to past forms, the dominate tense in narrative, will help accomplish the transformation.

When narrative transformation is not feasible, however, designers of instruction should consider ways to implement the other research finding noted above—that recognizing the salient features of a schema helps readers comprehend and recall text. Meyer (1975; 1982) has identified top-level structures in explanatory prose (in addition to narration) that have an impact on readers' comprehension in that these structures signal the schema and thus cue the reader's memory search:

- antecedent/consequent: presents causal relationships
- comparison: presents two opposing views
  - alternative: gives equal weight to both sides
  - adversative: favors one side
Her work suggests that the easiest passages to recall are first the
adversative comparisons, and then the causal relationships. Following
these structures in ease of recall is problem-solution, with description
being the most difficult.

These structures can be signaled in the text, thus making the schema
more salient and the text more readable. For example, words such as
"nevertheless," "still," "all the same" signal the reader to call up the
schema for adversative comparison. Headings, too, have been shown to
evoke the appropriate schema (Huckin, 1983). The repeated use of signals
and headings keeps the schema activated in short-term memory.

Another global readability principle is predicated on the nature of
short-term memory (Huckin, 1983). Informational content can be kept
activated in short-term memory by text that uses the grammatical subject
to refer to "old" information, the topic that is the focus of the text.
Research has shown that texts comprised primarily of sentences with an
old-new construction are easier to read and remember (e.g., Haviland and
Clark, 1974; Vande Kopple, 1982). According to theory and research on
information focus, we divide a sentence into its given, or old, and its
new information when we read it. We perceive the old information as a
pointer to a direct antecedent in memory, and we search for it. When we
find it, we attach the new information to it. If we cannot find an
antecedent, we either construct an antecedent or view all the information as new, adding a node to our memory (Haviland and Clark, 1974). Consequently, sentences that contain first the old information (usually in the subject) and then the new information (usually in the predicate) are easier to process. For example, the second sentence is easier to process in B than in A below because the subject is not new information; it has an antecedent in memory:

A. Ed was given lots of things for his birthday. The alligator was his favorite present.

B. Ed was given an alligator for his birthday. The alligator was his favorite present. (Haviland and Clark, 1974, p. 514).

How this principle works over several sentences is demonstrated in the example below in which "old" information recurs in the subject slot:

Friends are important.
My two best friends are John and Mary.
Mary gave me a book that she brought on her trip.
This book is my favorite gift.
Other gifts have also pleased me.

Another relevant global feature is the parallel presentation of elements/items/points, the importance of which is supported by research on surface structures (Fodor, et al, 1974). To present elements in a parallel fashion, we construct equivalent segments of texts, sentences, and parts of sentences in the same pattern. Then the equivalence is built in; it does not have to be constructed by the reader. This principle is exemplified by a simple application—the wording of two contrasting sets of screen commands:
Finally, global readability is enhanced in texts that give an overview of the main ideas. An overview familiarizes readers with the key content in the text and helps them perceive the text's overall schema. Many studies have verified this feature as a principle to implement when developing or conducting instruction. A relatively recent and brief review of these studies can be found in Luiten, Ames, and Ackerson (1980).

SENTENCE-LEVEL PRINCIPLES

A number of features that enhance the readability of syntactic structures have been identified. The discussion of these features is brief because they are simpler and more familiar than are the global features above.

Verbal Style

Sentences constructed in the verbal style are easier to read than sentences constructed in the nominal style (e.g., Davison, et al, 1981; Coleman, 1964). In the pairs of sentences below, sentence A, in the nominal style, is harder to read than sentence B, in the verbal style:

A. Their discussion concerned a tax cut.
B. They discussed a tax cut.

A. There was precision in the preparation of the data.
B. They prepared the data precisely.

A. The police conducted an investigation into the matter.
B. The police investigated the matter.

(adapted from Williams, 1981, 12-16).
Hake and Williams (1981), who have studied the nominal-versus-verbal issue extensively, approached this topic from an unusual perspective in one series of experiments. In an attempt to avoid confounding the issue with other factors in comprehension, such as readers' varying vocabulary knowledge or their personal strategies for approaching a text, Hake and Williams measured the dexterity of typists who had different levels of expertise: high school, college, secretarial college, and professional typists. Typists were timed as they produced two texts that had the same content and length, but different style—one was written in nominal style and the other in verbal style. All four groups typed the verbal passage faster, as much as 20% for the high school typists. The typists also had fewer errors when they typed the verbal passages—up to 43% fewer for the secretarial-school typists.

Noun Strings

Noun strings are easier to process when they are "untied" (e.g., Charrow and Charrow, 1978; Fodor, Bever, and Garrett, 1974; Holland, 1981; Gleitman & Gleitman, 1970). Noun strings are difficult to process because they contain only content words. Missing from the string are the function words that show the relationships among the content words themselves and the relationships of the words to the rest of the syntactic structure. In the examples below, A presents a noun string that is "untied" in B:

A. computer conference report information
B. information for the report on the computer conference

A. automobile polish users' brochure
B. brochure for users of automobile polish
A. lunch machine coin slot
B. slot for coins in the lunch machine

**Conditionals**

Sentences with conditionals (if...then, or...or, and...and) are hard to comprehend when the conditionals appear in multiples. Because conditions are easier when they are separated (e.g., Holland and Rose, 1981; Wright and Reid, 1973), they should be incorporated into sets of sentences or enumerated points. Conditions are combined in A but separated in B below:

A. If you are a senior and if you have completed your graduation requirements, you may participate in the class field trip, but not if your grade average is below C.

B. You may participate in the class field trip if
   - you are a senior, and
   - you have completed your graduation requirements.
   - and
   - your grade average is at least a C.

**Right Branching**

Studies have shown that sentences with modifiers that are embedded in the center and sentences with modifiers that branch to the left are more difficult than sentences that branch to the right (e.g., Clark and Clark, 1968; Davison, et al, 1981; Stolz, 1967; Schwartz, Sparkman, & Deese, 1972). Left branching and center embedding violate the normal and thus most frequent pattern of English sentences, Subject-Verb-Object and Modifiers, so they are more difficult to read. Thus they should be used sparingly and purposefully (e.g., to provide variety, to emphasize an element of the content).
Left branching is exemplified in A; right branching is exemplified in B:

A. much more easily understood facts
B. facts understood much more easily.

The center embedding in A below is contrasted with right branching in B.

A. The boys, when they had finished their work, went downtown to the movies.
B. The boys went downtown to the movies when they had finished their work.

Negatives

Affirmative statements are easier to read than negative statements, particularly when the statements are comprised of multiple negatives (Holland, 1981; Davison, et al, 1981; Clark and Clark, 1977). In negative statements, the reader must process the content of the statement and then negate it. This double processing slows down comprehension.

Multiple negatives are underlined in sentence A of each pair below, while B is recast in the affirmative:

A. He did not fail the exam.
B. He passed the exam.

A. No paper will be accepted unless the teacher reads it first.
B. A paper will be accepted if the teacher reads it first.

This principle applies to words as well, as in "remember/forget" (Holland, 1981).
Active Voice

Statements in the active voice are generally more readable than those in the passive voice (e.g., Davison, et al, 1981; Layton & Simpson, 1975). A major reason is that in active voice the grammatic categories of subject-verb-object, the natural order of English, parallel the semantic categories of agent-action-goal. These categories are noted in the sentences that are compared below, again with A in the less readable passive and B in the more readable active:

A. a ticket was purchased by the young man.
   subject verb object
   goal action agent

B. The young man purchased the ticket.
   subject verb object
   agent action goal

An exception occurs for this passive-active principle when the goal is the focus of the information. Then passive is appropriate because the goal/focus appears first in sentence order. This exception is displayed below with an example of a television programming announcement:

A. You will not see Tarzan today because of the following goal presentation.

B. Tarzan will not be seen today because of the following goal presentation. (Davidson, et al, 1981)

Relative Pronouns

Readability is often affected, particularly for young readers, when relative pronouns (e.g., which, who, that) are omitted (e.g., Holland, 1981; Charrow & Charrow, 1978). The reader must then infer the
connections between the modifier and the modified. This feature is exemplified by sentence A, which omits the pronoun, and sentence B, which includes it:

A. The teacher wants the report written by the best student.
B. The teacher wants the report that was written by the best student.

A. The school newspaper lists students awarded prizes in the contest.
B. The school newspaper lists students who were awarded prizes in the contest.

WORDS

Several word-level considerations are relevant to readability. The most obvious, of course, is the grade level of the words in the text. Various listings are available, such as the Harris and Jacobson word list (1972). Designs should first choose an appropriate word list. Then they should set a criterion for the grade level of the words to be used in the text. Limiting words to those listed at least one grade level below the level of the student population is one standard that some designers use.

Factors other than grade level affect the readability of words. Length is one factor, although research results are mixed. Studies have shown that shorter words are learned more quickly, but shorter words do not affect speed of recognition when the text requires procedures to be learned (e.g., Gibson, Bishop, Schiff, & Smith, 1964). Studies have also shown that words that appear more frequently are easier to read (e.g., Carroll, 1971).

Another relevant word-level feature is the concreteness of a word; for example, "money" is more concrete than "finances." Research has
shown that concreteness is more important to readability than is either the length or frequency of a word (e.g., Paivio, 1969; Cox, 1978; Katz & Denny, 1977). Related to this finding is the principle that complex words (e.g., wander) may aid understanding more than simple words (e.g., move) when the complex words are more specific (e.g., Kintsch, 1974).

GRAPHICS

To display text on computer screens, designers must also consider screen design as it affects the readability of text. Screen design is a relatively new research area, so the principles currently promulgated may soon be outdated by completed research that is still in press.

Alfred Bork (1981) posits a number of screen design factors that he believes can enhance readability. Those that are supported by research are noted here. For example, designers should use large amounts of blank space (e.g., Holland, 1981), and lines should be relatively short in relation to the size of the screen (Holland, 1981). Right margins should be ragged rather than justified because each line then looks different, thus making it easier for the reader to distinguish the line being read from the lines above and below it. Additionally, lines of text should not be broken in the middle of natural phrasal units, e.g., elements of a noun phrase, elements of a prepositional phrase (Snow and Coots, 1981). Although some research finds no difference between ragged and justified right margins, Gregory and Poulton (1970) found that poor readers have more difficulty with justified text. Research also suggests that the space between the lines in screen displays should be at least one-half the size of the characters themselves, (VanCott and Kinkade, 1972).
Displays that employ listings and headings also enhance readability (e.g., Holland, 1981), particularly when headings are cast in upper case and text in lower case (VanCott & Kinkade, 1972). All upper case should not be used for text other than headings (e.g., Poulton & Brown, 1968).

Research has also shown that highlighting important information enhances readability as measured by tests of recall (e.g., Crouse & Idstein, 1972). However, overusing highlighting techniques can have a deleterious effect on readability (e.g., Glynn & DiVesta, 1979).

Other screen design factors are important (e.g., the use of illustrations and graphics), but they do not pertain directly to the readability of the prose text, so they are not included here. However, employing the principles of readability described above will enhance students' comprehension of their texts, whether that text is displayed in books or on screens, and thus help them learn from what they read.
References


Kintsch, W., & Keenan, J. Reading rate and retention as a function of the number of propositions in the base structure of sentences. *Cognitive Psychology*, 1973, 5, 257-274.


