A retrospective look at predicting and producing readable computer documentation.

Readable Computer Documentation

George R. Klare
Distinguished Professor Emeritus of Psychology and
sometime Dean of the College of Arts and Sciences
Ohio University
8800 Johnson Road, #108
The Plains, OH 45780-1275
klare@oak.cats.ohiou.edu

Abstract

A retrospective look shows earlier advice still relevant to both predicting and producing readable writing. For prediction, refined readability formulas with stronger criterion passages and updated familiar-word lists have appeared, although the computerization of readability tests sometimes encourages misapplying or misinterpreting them when screening text. For production, attention to sentence construction, word characteristics, and information density remains relevant to both drafting and revising computer documentation for readability, especially since reading speed and reader preference often interact with comprehension in practical settings.

I.7.5 *Document analysis—human factors*

Keywords: *readability formulas, comprehension, document drafting*

Bob Waite e-mailed me March 30, 1999, to ask if the *Journal* might reprint chapter one of *The Measurement of Readability* (Klare, 1963) in this August, 2000, issue. He said if I agreed he would solicit commentaries from respected scholars and publish them along with the chapter, then ask for my response. Various thoughts crossed my mind while pondering this request. Not having read the chapter for quite some time, my first question concerned exactly what was said so long ago and whether I would still wish to say it that way. Something more recent seemed more suitable. But that would have run counter to his request to "feature a classic book, article, or report that has lasting value" for the profession.

This flattering characterization led me to re-read the chapter, "Useful Information for Communicators." Though the title now sounds a bit presumptuous to me, little appeared to need change in the first three sections, "Readability for the Reader," "The Writer's Purpose," and "Principles of Writing Readably." The commentators on the chapter might, it seemed to me, still generally agree with what was said then, which is the case now that their comments have arrived.

However, the final three sections of the chapter, "How to Use a Readability Formula," "Which Formula to Use," and "Limitations of Formulas," not only are dated, but can wave red flags in front of some writers. This has again been the case; all four commentators have expressed some degree of skepticism about the applicability of readability formulas to computer documentation. The comments of Hargis and Zibell question if, where, and when readability measurement might be applied to various kinds of computer documentation. Schriver takes a dismissive attitude toward readability formulas, feeling this approach is not only dated but outdated, and Redish takes a scorched-earth attitude toward almost all

ACM Journal of Computer Documentation, 2000,24:148-168
 ISSN 1527-6805

aspects of readability measurement.

Such feelings did not surprise me; I expected that criticisms might appear in the commentaries. After all, published criticisms of readability formulas date back at least to Moore's prediction of readers "recoiling from reading" rewritten classics (Moore, 1935) on to Duffy's "Readability Formulas:What's the Use?" (Duffy, 1985) and to Redish's own paper with Jack Selzer (Redish and Selzer, 1985). Well over a dozen others appeared in the years between. As the commentators have noted, my section on limitations in the chapter acknowledged many such criticisms in 1963 and agreed on their appropriateness.

Criticisms can be helpful. One, for example (Stevens and Stone, 1947), led to the improvement of Rudolf Flesch's original formula (Flesch, 1943, 1948). The aggressiveness of certain criticisms, however, reminds me of the psychologist Joel Greenspan's humorous comparison of ways to get attention for one's work (i.e., to get published) in the physical and social sciences. In the physical sciences, he said, one must stand on the shoulders of those who have gone before; in the social sciences, one can step in the faces of those who have gone before.

Bob Waite must have expected some misgivings on my part in responding to commentators: he promised me "the final word." He could not promise me the last word on this matter, but the final word in this issue of the *Journal* added enough amusing impetus for me to agree to his request. Before beginning my response, however, a few observations about my background for responding seem in order.

Background for a Response

Let me first acknowledge that I have never written about computer documentation specifically; technical writing yes, but computer documentation no. I was not even aware that some writers/designers of computer documents might have been using my book. Consequently, my readings of the four papers have been interesting and have filled gaps in my background. But does that qualify me to write about computer documentation without having had first-hand experience?

The answer must come from the readers of my response, of course. I have, at least, had first-hand experience with the need for clarity in computer documents. The beginning came in 1968 when serving for a year as a member of the staff of Lawrence Stolurow's pioneering Computer-Aided Instruction Laboratory at Harvard University. At that time I also met Edmund Berkeley, another computer pioneer, journal editor, and author of *Giant Brains, or Machines That Think* (Berkeley, 1949). Though his title might now seem hyperbolic, his approach to computer documentation was very down-to-earth (see Berkeley, 1967). One of his pet peeves was poor writing about computers, and his favorite term for it was COIK, standing for "Clear Only If Known." That, of course, is still a criticism of some computer documentation today, as my later comments will show. (As an aside, Berkeley prohibited smoking in his shop, not only because it could be unhealthy to humans but because a burning ash could prove unhealthy to the paper tape that ran his computer in those days.)

Another evidence for me of the need for improved computer documentation came from first owning and operating a personal computer a dozen or so years later. Much of the chat among friends who also had computers involved the inadequacy of the associated documents, and this is still a topic today. This concern for improved clarity of computer documents must have become widespread, since the "Dummies" and "Idiots" series of how-to guides soon appeared. The books were obviously very popular, despite their off-putting titles, as shown by a recent news release, "MIT receives pledge of $350 million to create institute of brain research" (March 10, 2000). The donors of the gift were Patrick and Lore McGovern, the founders of International Data Group and publishers of the Dummies series.

Writers/designers of computer documents are, of course, always on the lookout for ways

to clarify what they have to say. The four commentators in this issue, as well as the existence of this *Journal*, attest to that. I empathize with this concern, since computer documentation is so omnipresent and since users of some personal computer documents range from the nerd to the scared. How can I say anything that might be useful about readability and computer documentation? Are there any things other than the general principles of readability that can be useful? Do readability formulas have any place in computer documentation? If so, what?

My intent, by presenting research-based information, is for writers of computer documents to decide for themselves if, where, and when they might profitably use readability principles and formulas.

Approach in This Response

My best approach seemed to be maintaining as balanced and research-based an attitude as possible. I have tried to follow such an approach over the years and will again try to lay out, in an even-handed manner, useful information about readability. Thus, contrary to Redish's statement beginning "Klare, Flesch, Gunning and all the other developers of readability formulas...," I have never developed a formula. "Klare Formula" does have a rather appropriate sound (to me, anyway), but no, there is no Klare formula. A balanced approach to reviewing the literature has demanded no less. (Perhaps that approach has been carried *too* far: I get no royalties from sales of the 1963 book. I have not regretted that, however, since the National Project in Agricultural Communication provided funds for the research leading to the book, support which was welcome at that time.)

My intent, by presenting research-based information, is for writers of computer documents to decide for themselves if, where, and when they might profitably use readability principles and formulas. Yet I know that whatever might be said in my paper is not likely to mollify the clenched-teeth attitude some writers display when the term "readability formula" comes up. The term "readability measure" might be substituted since some devices like the popular Fry Graph (Fry, 1977) are not formulas. But that would not fool anyone. Similarly, I might use the term "clarity" rather than readability, as my wife and I did in naming a writing partnership some years ago (we resisted the suggestion of spelling the word with a "K"). But again, that would not really change the picture, so I will generally use the traditional terms in referring to the research.

Some document developers prefer the term *design* rather than *writing*, since it emphasizes that there are aspects other than writing in producing documents. The compound noun *document writer/designer* seems unnecessarily cumbersome, so where appropriate I will use a new term, *documentor* (note the "or" ending). There is always danger in employing new terminology, as documentors well know, since that is one of the complaints they face from document users. Nevertheless, the term is appealing not only because it shortens the compound noun but also because the "mentor" fragment emphasizes the instructional aspect of computer documents.

As noted above, the four commentaries came packed with useful information about issues in computer documentation and pointed comments about readability. I have learned a great deal about the development of computer documents, particularly different kinds of clarity and usability problems for the different kinds of documents in which they appear. Of special note for me were Hargis' "Levels of Readability" and Zibell's analysis of web design. While Redish and Schriver mostly find fault

with readability formulas, they argue cogently for usability testing as the best approach to developing documents.

Each writer organized her information in somewhat different fashion, which created a problem for organizing my response. The temptation to comment on the individual points in their different contexts as they came up would have meant a fragmentary, hop-scotch response. This concern led me to feel that my overall division used before, between the prediction and the production of readable writing, would be most suitable. If readers wish, they can in their own minds change the terminology to suit.

The prediction issue, one more easily covered than the complex and difficult production issue, will be considered first. A research-based approach seems appropriate in responding to what I consider misapprehensions about, and misinterpretations of, readability formulas and their proper use. Could my writing have suggested readability formulas can be applied appropriately to all kinds of writing for all kinds of purposes to all kinds of readers? Redish's comment from her paper with Selzer (Redish and Selzer, 1985, p. 49) seems to imply that. If so, let me correct that misapprehension in the following section.

Special attention in the section will also be given to newer work in the field and to little-known but relevant older publications. These will be described briefly. The many publications on readability prevent me from describing all the studies that support my statements here or that deal with the blanket condemnations by Redish and Schriver of formula reliability and validity. Readers must decide on the appropriateness of their criticisms on the basis of the published research. Recent work by Jeanne Chall (Chall and Dale, 1995) and by Edward Fry (Fry,1989) present the other side of this issue. My extensive review of the literature in the *Handbook of Reading Research* (Klare, 1984) provides comments and references for both sides of this matter.

Prediction of Readable Writing

The prediction of readable writing has traditionally rested upon use of readability formulas and the scores they yield. Well over one hundred formulas have been published for different purposes, some carefully developed and some simply based on existing formulas. Many hundreds of studies of readability have been published using or misusing the formulas. These publications cover such technical matters as criterion data, sampling, analyst reliability, and internal and external validity as well as applications of formulas to varied bodies of text. Since all four commentators have, as noted, expressed concern with the formulas and their application, responses of agreement or disagreement on specific points are in order. But first let me summarize briefly what readability formulas are and are not, then present my views. Using this background, I can more easily refer to comments in the four papers.

Typical readability formulas are statistical regression equations, not mathematical identities, and do not reach that level of precision. The readability scale, unlike the temperature scale, does not have a zero point and cannot therefore represent absolute amounts or degrees of readability. Like most psychological and educational scales (e.g., IQ), the readability scale reliably represents only "more than" or "less than" amounts or degrees. (Even the temperature scale is modified occasionally with a wind-chill correction to represent how the temperature feels to humans.) Consequently, a readability formula is properly considered a screening device that provides probability scores for a quality in text not otherwise easily represented.

With that information in mind, let me summarize my view of readability formulas and their application in the following ways. Most of these qualifications appear in a listing in the1984 review (p. 730) intended for educators and reflecting the formulas' educational heritage; this listing is modified somewhat here for computer documentation.

The better readability formulas:

- Combine weighted counts of language indices (most often, word and sentence variables) to estimate semantic and syntactic causes of difficulty in text.
- Touch only upon what might be termed "difficulty of style," and do not take account of format, organization, and other such text characteristics.
- Apply only to what has been called "connected discourse," and are therefore not applicable to sentence fragments, word lists, tables, etc.
- Apply properly only to entire bodies of text or at least a number of large, representative samples (most often, 100 or so words in length).
- Yield an imperfect indicator score on a scale of reading difficulty (usually grade levels), with such scores considered as probability statements. .
- Apply best, if a wide-range formula, where difficulty of style overshadows difficulty of content (i.e., roughly fourth grade to beginning college level).
- Apply most appropriately to material for children or for adults with limited educational aptitude or background.
- Function best as screening devices only, and do not cover such potential contributions to comprehension as tables, graphic presentations, etc.
- Should be interpreted in light of such reader characteristics as motivation, background knowledge, interest, and purpose in reading.
- Cannot be used as formulas for writing style, since changes in their index variables do not produce corresponding changes in reader comprehension.
- Should be used in conjunction with other approaches to evaluation of writing, such as use of judges, cloze procedure, and usability testing.

Comments on the Commentaries

Schriver, in the introductory section of her paper, says that the "call for new and better formulas...went largely unheeded" from the late 1970s through the mid-1980s onward as attacks proliferated. The attacks did increase for a time, and the number of articles on readability declined somewhat, but there is a highly significant exception to her statement regarding new formulas. A new "Chall-Dale" formula (Chall and Dale, 1995) based on the earlier Dale-Chall formula (Dale and Chall, 1948) appeared. Over the years since it was published, the original Dale-Chall formula proved the most predictive of the wide-range readability formulas. Despite widespread use, its early appearance and the many details in the word count limited its accuracy in applications. The two authors began to develop the new formula some years ago, but Edgar Dale's death intervened and the work was completed only recently by Jeanne Chall. This formula could have been called a revision, since it builds on word and sentence counts like the earlier one, but it does make significant changes. Since the new work is not referred to in Schriver's or any of the other papers, yet touches on comments in them, brief information about the changes is in order. The new formula, compared to the original, has the following characteristics:

- Stronger sets of criterion passages have been used in development, with correlations ranging from .64 on multiple-choice comprehension scores to .92 on cloze comprehension scores. (For readers unfamiliar with cloze procedure, see Klare, Sinaiko, and Stolurow, 1972, for example.)
- The original list of 3,000 familiar words has been updated, based on extensive testing of word meanings known by students from fourth-grade level to college level by Dale and O'Rourke (1976, 1981).
- Some of the rules for the word and sentence counts have been modified.
- Details of application have been simplified for greater efficiency.

- Readability scores in terms of both grade levels and cloze percentages have been provided.

Chall and Dale's research, plus my list of qualifications above, answer certain concerns raised in the four commentaries. A word first about the assertion of Duffy (1985), echoed by Redish, that formulas are suspect because they use only a 50% correct criterion for 50% of tested subjects for assignment to a grade level. Criticisms of using school grades as levels of difficulty aside (and there are some), the statement can be misleading. The 50% criterion originally proved best for statistical development purposes, but not for assignment of grade levels. Most formulas, and certainly the better ones, correct that figure by assigning grade levels for 70 or 75% correct, as noted by Hargis.

McLaughlin (1969) actually used a criterion of 100% correct in developing his SMOG formula, which resulted in scores about two grade levels higher than typical formulas using a 75% criterion. This 100% level could not be reliably expected in application, since comprehension tests themselves are not 100% reliable. No argument can be made that formulas predict such precise grade levels of all text. Only screening use is therefore advisable and, as Hargis points out, if a comprehension level as high as 90% to 100% is desired by documentors, a lower grade-level readability target can be used. Futhermore, Chall and Dale provide cloze scores as an alternative metric for those who do not wish to use grade levels.

Redish, in the related statement with Selzer alluded to above, says, "the underlying assumption of readability formulas—that any text for any reader for any purpose can be measured with the same formula—does not mesh with our current understanding of how people read and understand." This argument runs afoul of several of the qualifications listed above. Some formulas have been developed for the early grades (up to four) or just for college-level readers, for example, rather than for all readers. Some have been developed for arithmetic text, or listening, or human interest, or for languages other than English, etc. But even the wide-range formulas for English text are again best viewed only as screening devices. This limitation, plus the need mentioned above for analyzing either entire documents or at least extensive samples, voids the example Redish presents in the two four-word sentences with the words "wave" and "waive." Her assertion that "these two sentences have identical scores," while itself inappropriate, deserves further comment below.

There is no doubt that readability formulas have sometimes been misapplied or the resulting scores misinterpreted. Redish provides an example herself when she says that "readability formulas will say you have long sentences" when you use bulleted lists. But formulas cannot be applied appropriately to lists, as noted above. Schriver argues that you "doctor the text" when you remove such lists and might as well therefore not use a formula. But bulleted and numbered lists that do not use complete sentences constitute only a minor part of most text; there is no reason therefore why they cannot be omitted in formula applications while evaluating the bulk of the text. Many writers (myself included) favor such lists when they seem appropriate, but do we know what effect they have on readers'comprehension compared to their straight-text counterparts? Research is needed here and on many other assertions about writing techniques. The results might well prove surprising, but discussion of such research will be postponed until later in the "Production" section.

Redish provides a good example of another unfortunate kind of misuse. She says, "many groups that use the Dale-Chall formula improve their scores by arbitrarily adding to the list of acceptable words those words that they *believe* their audiences know." She gives the word "enter" as an example. As she points out, a fourth-grader's knowledge of the word's meaning is not a good indicator of the word's meaning when used in a tax form. No one could disagree with that.

Adding words appropriate to a special context has been tried, particularly with science terms

(Brown, 1965), and the effect of the change in predictive power actually tested on readers (Holmquist, 1968). This kind of revision can have only a very limited effect upon predictive power, as Holmquist found, because the Dale-Chall formula(s) only provide difficulty-level scores based on the proportion of familiar words in text. They only predict levels of overall style difficulty, and were never meant to be arbitrarily modified in the above fashion. However, the criticism that the list of words in the original Dale-Chall formula was developed long ago is relevant; language changes and grows. Chall and Dale have, as indicated, noted this and modified the word list for their formula. The crucial point is that the words they have added or deleted were based on research, not on a whim.

The different meanings for Redish's two different words "wave" and "waive" are, of course, obvious. But "wave" itself has different meanings, as does "waive." Edgar Dale devoted an almost lifelong research effort to documenting the difficulty of the different meanings of polysemous words, culminating in the *The Living Word Vocabulary* (Dale and O'Rourke, 1976, 1981). Looking up Redish's two words indicates that the meanings of "wave" as a ripple on water, a curl in hair, or a hand signal are familiar to most readers with a fourth-grade education. However, "waive" as putting off or postponing, or as giving up a right, are familiar only to the majority of readers with a twelfth- or thirteenth-grade education. Even the meanings of "wave" as a line of attacking soldiers or as a motion of vibrating particles are, Dale and O'Rourke found, familiar only to the majority of readers with a twelfth-grade education.

Screening Use of Readability Formulas

Ironically, the widespread availability of personal computers and of readability formula programs for them have contributed to the misapplications and misinterpretations of readability measures. It is easier than ever to apply a formula to a text while in the writing process and make an assertion about its difficulty; too easy in the view of many (including me). Perhaps several colleagues and I share some culpability for this since we (Klare, Rowe, St. John, and Stolurow, 1969) developed one of the early computer programs for applying a readability formula. I mention this development for two reasons.

First, few others used our program, since application required a box and a half of IBM punched cards. We had found it difficult to develop an accurate algorithm for counting syllables and finally had to add a look-up list for increased accuracy. Compare, for example, the syllables in the similar-appearing words "real" and "read." Many programs today estimate syllables from word lengths in letters, a simpler tasks for computers. This expedient adds error variance to the error variance inherent in using a syllable count as an index of semantic difficulty.

Second, I mention the above program in order to describe its use in a little-known screening study. A colleague and I (Klare and Smart, 1973) analyzed a large-scale sampling (over 132,000 words) in 17 sets of United States Armed Forces Institute printed instructional materials. We found high correlations (rank-order, .87, and Pearson, .76) between the readability scores and the probability that students would finish their courses, with length of courses held constant. We had to include one qualification: that the students had turned in the first lesson. Otherwise, the reading difficulty might not have evidenced itself. Besides, we knew that at least some students enrolled in the courses primarily in the hope they could sell the books they were given.

This screening was difficult in the sense that formula application required that all printed text materials be word-processed again, yet it uncovered an interesting relationship that deserves greater attention later. Screening applications usually need not be so extensive as this to provide useful information. And, as noted, the additional word-processing effort need no longer be made when writing is underway directly by computer, as most documentation is these days.

Hargis makes some valuable points along the

line of applications in her paper, particularly in terms of her interesting categorization of levels of readability. A great deal of research has been directed at the "readable as comprehensible" level, as several reviews of the literature show. She points out, however, that readability analyses might well contribute more at the level of "readable as translatable." A screening look using readability formulas is clearly possible.

Hargis is not referring to readability of computer translations, of course, but a colleague of mine and I (Sinaiko and Klare, 1971) did investigate this area some years ago. The measure mentioned above that we used, cloze procedure, is worth looking into since it has applications to translations more generally. Robert Kern has also noted (e-mail reply, April 23, 1999) that a screening application of a traditional readability formula by students to their own material can turn them on to principles of readable writing. He reports this reaction for his own students in Bangladesh where translations were involved.

To try screening on a computer document, the suggestions for users presented in my review article (Klare, 1984, p. 730) can be of help. Since they are available in published form, I will not repeat them here except to note a suggestion that users avoid relying on formula predictions alone when this can be avoided. Instead, expert judges were recommended; they can evaluate aspects of writing that formulas do not cover, and be sure formulas have not been misused unintentionally or intentionally by one of Schriver's "ornery" writers. Writing today for documentors, I would suggest that a usability test follow where appropriate.

Does applying a formula mean that Kern's students must reach the "seducible moment" of Schriver and succumb to the temptation to write to formula? He did not find that, but it can happen, just as the IRS documentors were apparently tempted to doctor the Dale-Chall formula to reach some desired readability score. In this regard, incidentally, *The Living Word Vocabulary* estimates that the meaning of "enter" as "put in record" is familiar to most eighth-graders. In other words, the appropriate sense is more readily known than might be expected. One can only hope IRS documentors become aware of this and take it into consideration in their analyses.

The value of *The Living Word Vocabulary* for writing needs mention, but that discussion fits better in the following section on producing readable writing. A final point about readability formulas needs emphasis before moving on: they can be applied appropriately *only* with the limitations listed above. Documentors who have material where formulas might be applied as screening devices and who wish to try one will find the procedure has recently been simplified. A newly published computer program accurately covers, in my view, the three best readability formulas. This program (Readability Master 2000, 1997) is available in both Windows and Power Macintosh formats and applies the following formulas:

- The new wide-range Chall-Dale formula described above (Chall and Dale, 1995) is for text primarily at or above fourth-grade level in difficulty.
- The wide-range Fry method noted earlier (Fry, 1977), also for text primarily at or above fourth-grade level in difficulty, is normally available as a monograph. It uses a different index variable for semantic difficulty (syllable counts) than the Chall-Dale formula, and thus can provide a good check on Chall-Dale scores.
- The Spache formula (Spache, 1974) is for use only with primary grade-level material.

Production of Readable Writing

Producing writing that is readable to a particular reader is a complex and difficult matter. Documentors may or may not wish to try a readability formula for screening a document, but would certainly not wish to write to a formula. As noted earlier, formulas use only index variables for the more complex causal variables of semantic and syntactic difficulty. Furthermore, they should not be

used as an alternative to usability testing or other evaluative measures such as judges or cloze procedure. Usability testing can follow screening since the two are compatible, and the value of usability testing in the development of documents is clear. Redish provides half a dozen good recent sources for usability testers. Hargis mentions protocol analysis as a useful way to learn how people use software, which is again clear.

My limited experience with protocol analysis (Schumacher, Klare, Cronin and Moses, 1984, and Schumacher, Scott, Klare, Cronin, and Lambert, 1989) suggests it has a great deal to offer for research on the cognitive processes during complex mental tasks. However, a full-scale protocol analysis appears too time-consuming and labor-intensive for most uses other than research. It cannot provide the relatively quick, inexpensive, and easy evaluation that Redish credits to usability testing. Consequently, I doubt that many document developers use protocol analysis on a routine basis. This is not to say that it has little value when it comes to understanding the writing process. The publications of Flower, Hayes, and their associates have stimulated a great deal of interest in a cognitive process theory of writing (Flower and Hayes, 1981, Flower, 1985, and Flower, Hayes, Carey, Schriver and Stratman, 1986).

My experience with usability testing has been more limited but goes back farther, prior to its widespread use with computer documentation. In 1976, an internal Bell Labs report of mine reviewed several methods of technical writing, one of which was the Heath Company's approach. The company needed assurance that purchasers could successfully assemble their highly detailed Heathkits for high-fidelity sound components and television sets. Employees of the company were therefore offered free kits on condition they report any problems they had in assembly. The approach must have been of value since Heathkits were widely used, but it was not foolproof.

As an example, let me mention a talk with an electronics engineer in the Bell System who had assembled a Heathkit. He told me he put together either the amplifier or the tuner (which one escapes me), and his wife, who had no background knowledge of electronics, the other. When they had finished their tasks, his wife's unit worked but his did not. He said he had skipped some of the instructional material, feeling that he knew all that stuff already; his wife, he said, skipped nothing. As this anecdote suggests, a usability test of a document cannot guarantee successful application. Too often, some further instruction or even a bench-test proved necessary to achieve an operational Heathkit. More important, however, this example says something about problems in the testing and revising of documents more generally.

Revising Documents

What should a documentor do when a reader/user has trouble in usability testing? Redish mentions a number of good ways to correct problem documents, with the readability variables of style and word choice among them. However, she lets the matter rest by saying that usability testing tells the documentor "how to fix them" and also lists important aspects of documents that readability formulas do not cover. Where does that leave the documentor who wishes specific information on changes, or on original drafts for that matter, that can optimize reader comprehension?

Books on writing style can be of help, and documentors doubtless have their own favorites; grammar-checker programs can also be used (but probably seldom are). Psycholinguistic research can now provide some tested results on the effect of style changes, and Redish lists several references on linguistic aspects that make sentences difficult. My book *How to Write Readable English* (the insistent publisher's title, not mine) provides research information on six sentence constructions as well as six word characteristics that can lead to difficulty for readers (see Klare, 1985). More recent research could add to the list.

Documentors once again must have their favorite aids in the matter of word choice, and

synonym programs are now available in one form or another. A book that may have escaped some documentors' notice deserves mention. *The Synonym Finder* (Rodale, 1978, 1986) provides "over 1 million synonyms," far more than any program. While not all of the listed words are truly synonymous, the suggestions can help documentors choose the precise meanings they intend readers to get. Various books can help in making fine distinctions, but none can be of as much potential help, in my opinion, as *The Living Word Vocabulary*.

Edgar Dale felt disappointed during his lifetime that this volume was given so little attention. He did not argue that the compilation gives perfectly accurate indications of the educational ages when particular word meanings become familiar, but his data come much closer than guesses. For many writing purposes, his count is much more useful than the better known word-frequency counts. Assuming one knows the stage at which a reader understands the intended meaning of a word can be risky, as my son Roger Klare, a writer of science materials for children, has emphasized to me. He has found the *Vocabulary* invaluable in his work. The same might be said for documentors hoping to reach adult readers with limited educational backgrounds (more on this matter later).

At the same time, as stated in my paper "Assessing Readability " (Klare, 1974-75), "altering word and sentence length, of themselves, can provide no assurance of improving readability. How to achieve more readable writing is another and much more complex endeavor." This paper, named a Citation Classic (Current Contents, 1988) by the Institute for Scientific Information (but not cited by the commentators) makes a distinction between measuring and predicting readability. Readability formulas are intended for prediction, not measurement, of readable writing.

Redish endorses my statement that the word and sentence variables in readability formulas are only indices of the causal factors of semantic and syntactic difficulty, and refers to my thermometer analogy. My introduction of the analogy (Klare, 1979) goes on to indicate that there is no clear demarcation between indices and causes of difficulty, which Redish fails to mention. Some extremely high correlations that have been found between index variables and presumed causal variables give one pause. The following examples show such relationships:

- The number of *morphemes per 100 words* appears to be a cause of semantic difficulty. Yet the much simpler index of *syllables per 100 words* has been found to correlate with it to the extent of .95 (Coleman, 1971).
- The sum of *Yngve word depths per sentence* has been considered a cause of syntactic difficulty. Yet the number of *words per sentence* correlated .99 with this count in one study (Bormuth, 1966).

When correlations reach the values above, one is led to wonder whether the index variables might not be tapping causes. In fact, one seems to shade into the other at some point. For example, length of sentences alone, to say nothing of ambiguity, can sometimes be a problem, given human limitations on working memory. The high degree of relationship between the classic index variables used in readability formulas and causal factors examined in research on comprehension deserves further mention because it complicates the index-cause distinction.

As Jacques Barzun (1980) has put it, "Simple English is no one's mother tongue. It has to be worked for." Mental parsing is often enough to suggest to documentors how to change a sentence to make its meaning clearer. But that is not always an easy task. The linguist Derek Bickerton provides a good example in *Lingua ex Machina* (Calvin and Bickerton, 2000) in the headline "Spy Charges Dog Inspectors." Disambiguation is almost impossible unless one knows the context, which happened to be weapons inspections in Iraq.

An extreme attempt to avoid ambiguity can be found in what was called Caterpillar Fundamental English (1972,1976), which was included in my review of technical writing

for the aforementioned Bell Labs report. The Caterpillar Company was faced with either costly translations or training programs for users of their documents in the many countries where English was unfamiliar. The company tried providing a limited set of words required for the maintenance of particular equipment along with a set of operator words with fixed meanings that could be taught easily to users. For example, the word "right" could mean a direction or a hand designation but not an indication of correctness. A trip to Peoria to talk to technical writers at the company proved disappointing. The developer had been transferred to a different job, and technical writers hated writing under the constraints of the Caterpillar Fundamental English formula. This, in turn, often necessitated costly editing of material.

Documentors need not, in fact should not, write to a readability formula. Nor need they be limited to language when trying to express their meanings clearly. Achieving usable documents for members of an intended readership is, of course, much more complex and subtle than word and sentence changes alone, as Schriver's exhaustive *Dynamics in Document Design* (1997) makes clear. Does that mean improving readability through word and sentence changes has little, if any, effect on comprehension/retention?

Readability and Comprehension

Redish refers to my examination of 36 studies (Klare, 1976) that attempted to improve comprehension by employing changes in readability variables. As she points out, only about half yielded clearly positive results. The word "examination" is appropriate here because of my detailed review of the 36 studies. The strength of the relationship of readability to comprehension had puzzled me ever since an early study (Klare, Mabry, and Gustafson, 1955). My colleagues and I had found that more readable versions of technical training course material yielded significantly increased comprehension scores (which we called immediate retention in deference to the terminology of the time).

We also found that the more readable versions produced an even stronger effect on reading speed and reader preference than on comprehension. This differential has been noted in other studies before and since. The reason for the effect on reading speed/efficiency is fairly well understood. It has long been known (Zipf, 1949) that content words (nouns, verbs, adjectives and adverbs) as opposed to function words (articles, prepositions, conjunctions, interjections, etc.) become shortened with frequency of occurrence. *Horseless carriage* becoming *car* or *television* becoming *TV* are only dramatic examples of this process. Shorter words are part of the picture. More recent research has also shown that more frequent words are perceived faster even when length is controlled (see Richards and Heller, 1976, for example). The consistent effect of readability changes on reader speed/efficiency thus clearly rests on the use of more frequent/familiar words.

Where reader preference, or more generally "acceptability," is concerned, the effect is again clear. The early studies of Donald Murphy (Murphy, 1947a, 1947b) and others had already shown a relationship between readability and readership of newspapers and magazines. We found the same preference effect among trainees reading split-half technical lessons that permitted the ruling out of content preferences. The study mentioned earlier (Klare and Smart, 1973) found a similar relationship between readability and perseverance in terms of course completion. The underlying factor here appears to be what Zipf (1949) called "the principle of least effort." Informative reading is usually work and seldom fun, as Hargis's computer joke emphasizes; readers appreciate writing that makes their job easier.

Why were the few studies prior to our 1955 experiment not consistent in finding a readability-comprehension relationship? Presenting all the details of our study would take too much space here; the interested reader can consult the published article. Let me say only that we had the luxury of spending a

great deal of time in planning, developing, and carrying out all aspects of the experiment to make it as definitive as possible.

With that prologue, let me take up the examination of the 36 studies very briefly. Preparation for an invited address to the National Reading Conference in 1975 gave me an opportunity to examine as many studies of the readability-comprehension relationship as could be located. The subfindings have as interesting implications for computer documentation as the overall finding that only about half yielded positive results. In fact, the negative studies proved more valuable than the positive because they could tell more about why they failed than the positive studies could tell about why they succeeded. As a statistician might put it, one cannot prove the null hypothesis!

Doctoral dissertations were chosen for study where there was a choice in order to get at as many details of experimentation as possible. Because of length limitations, journal articles are sometimes as notable for what they omit as what they include. Dissertations, on the other hand, typically provide far more details because graduate students are able to hide little if anything from their mentors. Forty characteristics of the studies seemed promising, but such information was available on only 28 of them in all of the studies. These broke down into seven categories, which I summarized in a figure that related "reader performance" to the following major categories and their interactions:

- The test situation,
- Reader competence,
- Content of material,
- Readability level of material,
- Reader motivation.

The categories still seem appropriate to me except for the substitution of "prior knowledge" for "content of material" to represent current thinking. "Reader performance" refers not only to comprehension scores but also to reading speed/efficiency and reader preference/acceptability as mentioned above. Reading speed could not be evaluated in enough studies for a reliable finding, but the acceptability effect was found in a large enough number of studies where the design provided for it. In 10 of the 17 studies which yielded negative or mixed results on comprehension, for example, readers had been asked to pick out or indicate their preference for the more or less readable version. They consistently chose the more readable even though comprehension differences were not found to be significant.

Let me also say a word about the term "interacting with" that connected the various factors listed above. Each of the factors can have independent effects upon reader performance, but they may also interact with each other in their effects. The interactions uncovered by research have as much potential value to documentors as the independent effects. These relationships have not all been fully investigated, but some have been examined in published studies since the 1976 article. Of particular significance is a little-known experiment by one of my doctoral students (Entin and Klare, 1985). A description here quickly proved too detailed and too long for this paper, but the interested reader can turn to the article itself for details. Let me describe the experiment briefly, however, to provide a framework for the results of this and related experiments.

Why, when the relationship between readability and speed or acceptability is so clear-cut, is that to comprehension/retention not clear in the experimental studies? Was it because index variables were not powerful enough to produce an effect? Was it lack of skill on the part of the writers? Could it even have been intentional on the part of ornery writers? Such subjective factors were difficult to evaluate, but did not seem to be operating. Further study seemed desirable since the relationship of readability to comprehension was of primary interest to early formula developers, and is the focus of most experimental studies of the effect of readability. Furthermore, any value of readability modification for documentors is reduced if a comprehension/retention effect cannot be

expected.

The Entin-Klare study, taken in conjunction with the figure, provides some answers that help to explain this result. The study involved large samples of university freshmen and examined the four major factors listed above.

- Reader competence was represented by scores on the Nelson-Denny Reading Test (Nelson and Denny, 1960).
- Content of material was measured by prior knowledge of key terms from experimental passages (see, for example, Anderson and Freebody, 1981, 1983).
- Readability level of material was evaluated by passages at approximately 12th or 16th grade levels on the Flesch Reading Ease formula (Flesch, 1948).
- Reader motivation was measured through passages of high or low interest as determined by a modified reading interest survey (see Waples and Tyler, 1931).

Reader interest, prior knowledge, and readability level all produced significant increases in cloze comprehension scores in the experiment. Reading test scores were used only as a covariant to achieve comparable skill levels among groups studied; their relationship to comprehension has often been shown in the literature. That reader interest, prior knowledge, and readability level independently produced increased comprehension may not seem surprising. Logic suggests they should do so, although as noted for readability levels in the study of the 36 experiments, previous studies with these variables had not all produced positive results. Furthermore, the results counter the argument sometimes made that readability is basically only prior knowledge, since both of them produced independent effects. The possible interactions of these variables can provide added information.

Interactions Involving Readability

The interaction of motivation and readability has been studied by Fass and Schumacher (1978) and McLaughlin (1966). A monetary reward was used to produce a high level of motivation in the former study, and threat in the latter. The results demonstrated that, when motivation is high enough and readers have sufficient reading time, the effect of readability is attenuated: more readable passages no longer produce increased comprehension scores. In the Entin-Klare study, by contrast, more readable passages produced increased comprehension of both low-interest and high-interest passages.

Was this additive effect rather than the predicted interactive effect due to using reading interest as a motivating factor rather than monetary reward or threat? Perhaps, but documentors can seldom use either directly in practice anyway, so the question is academic. And, as Hargis suggests, adding interest to a computer document is so difficult that, in some cases at least, documentors do not even try. Or, as Zibell says, motivation is so low that "on the Web, everything seems to be voluntary." Surely there must be other ways to build reading interest into documents.

Rudolf Flesch, in his ill-fated Human Interest formula (Flesch, 1948), suggested that writers create interest in text by using personal words (e.g., "you") and personal sentences (those addressed directly to the reader). Little came of the formula itself, but writers have long known that a personal style, if used judiciously, can seem more interesting than an impersonal style. But a personal style can seem artificial if carried too far, just as sprinkling a reader's name through appeals for contributions or advertising come-ons now does. Readers once charmed by this "Dear John Smith" approach now realize that impersonal computers are doing the work on an impersonal basis.

Another approach for introducing interest is "analogic interest," building reader interest into the analogies, examples, and illustrations in documents. What do readers like to read about besides themselves (even when criticized!)? Four of the six high-interest topics used in the study may suggest useful applications. They were the nature of personality, why we forget, keeping physically fit, and how emotions help and harm us. The other two of the six, what happens when we sleep and attitudes about death, do not

seem so promising. Perhaps another approach based on readers' background knowledge might be possible.

This final variable examined in the Entin-Klare study, prior knowledge, again had a significant effect upon comprehension when considered independently. And once again, interactions were not found. The number of cases on which certain of the analyses were made was necessarily quite small, making reliable comparisons difficult. Subjects seemed unable to differentiate prior knowledge from interest. Very few subjects could be found who were low in prior knowledge about a passage but had high interest in that passage, and vice versa.

Tobias (1994) has made an extensive study of the relationships of interest, prior knowledge, and learning. He has come to much the same conclusions as were found in the Entin-Klare study. For example, he has found a substantial linear relationship between interest and prior knowledge. He also found that categories of high interest and low knowledge and low interest and high knowledge are likely to be transitory. He concluded that interest contributes to learning in these ways:

- Interest invokes deeper types of conprehension processes.
- Interest leads to greater use of imagery.
- Interest may stimulate a more emotional, more personal, and more extensive network of relevant associations than is invoked by prior knowledge.

The close relationship of prior knowledge and interest suggests that estimates of either variable might be used in developing documents that are comprehensible for certain readers but not overly redundant for others. The Open University in Britain has used a prior-knowledge approach by presenting a list of terms students are presumed to know before beginning a correspondence course. Such terms might be placed at the beginning of documents to let readers know whether their degree of background knowledge is sufficient to profit from reading a document.

The readability factor proved more powerful than either reading interest or prior knowledge in terms of the statistical variance covered in the experiment. Was that because the difference between the more and less readable versions was as great as four grade levels? Redish clearly feels that requiring readability levels 6.5 grades apart to get positive results, as found in my1976 review, seems unreasonable. She might well feel the same way about the four grade levels in the Entin-Klare experiment. Could positive results have been found with a smaller difference? It is too late to ask that question now, of course. However, two considerations about the strength of the readability effect that arose from the study are relevant to this question.

First, the two readability levels were either at the readers' tested competence level, 12th grade (entering college freshmen), or above, at 16th grade. In most experimental studies of the effect of readability, one of the levels was instead below the readers' competence level as well as one above, thus widening the distance apart. There is now evidence that material below readers' competence levels makes little if any contribution to comprehension compared to material at readers' levels. Thus, the difference of 6.5 grade levels might have been smaller in many of the positive studies if only comparisons of material at or above readers' levels had been employed. Chall supports this reasoning (Chall and Dale, 1995).

Second, the process of testing in itself increases motivation, and particularly when there is a time limit this tends to make it harder to achieve a significant difference. Our educational history in test-taking conditions us to react this way. My colleagues and I found this reduced the effect on comprehension for changes other than readability as well. We tested improved organization (use of paragraph headings) and stress (emphasis on critical terms) on comprehension of the same experimental passages with mixed results at best.

If it were possible to use unobtrusive measures of comprehension, smaller differences in readability levels than those used in experimentation

might well produce significant increases. But however strong the effect of improved levels of readability upon comprehension might be, it seems certain that persistence or other measures of acceptability would increase. Documentors might profitably consider this information about readability levels, since computer documents often look dishearteningly thick to prospective users.

Other studies of the relationship of readability to comprehension have been published since the 1976 paper. Mostly positive results have been reported by Ewing (1976), Kincaid and Gamble (1977), Greene (1979), and Swanson (1979). Negative results have been reported by Charrow and Charrow (1979) and Duffy and Kabance (1982). Redish describes the Charrow and Charrow study briefly, noting that in many cases their revisions got better comprehension scores but worse readability scores. I have not examined the details of any of these studies, but Robert Benson (1984-1985) reviewed the Charrow and Charrow study and found significant flaws in it. Perhaps a detailed examination of all such studies using the approach described in the 1976 study would again prove informative.

At any rate, the results of the Entin-Klare study and related research seem to provide some useful suggestions for the work of documentors. One finding in almost all such studies is that increased reading speed and acceptability can be expected, whether or not comprehension is improved by readability changes in text. We need tests of the effects of other kinds of text changes in the attempt to evaluate their worth in documentation as well as other kinds of applications. The next section discusses some observations on problems users have with computer documents, and also related questions.

Computer Document Problems: Some Observations and Questions

Under ideal conditions, this section would begin with the results of a carefully conducted survey of user problems with computer documents. I have neither made one nor am familiar with one. Here, however, are observations growing from my own experience with computer documents as well as discussions with a dozen persons who use personal computers. My daughter, Barbara Klare, a senior technical writer at Extensity, Inc., has been especially helpful. Documentors may already be aware of many of the problems mentioned here, so the emphasis will be on underlying issues regarding readers and their habits.

The number of criticisms of computer documents suggests either that a usability analysis has not been made or is an imperfect method. Or that an analysis has been misused by harried or ornery documentors, which would be easy enough to do. And these criticisms do not include the horrors of documents translated from another language, particularly a character-based language. The flip side may well occur when documentors use colloquial English that must be translated into some other language.

Usability testing appears to be an especially good approach for documentors where the instructions for a task are sequential, especially where one step must be completed successfully before being able to go on to another. Here, as Redish notes, the problems of trial users can be spotted by trained observers and the text corrected to eliminate future problems. Provided, of course, that the trial users constitute a representative sample of the population who actually use the documents.

One of the problems lies in the term "representative." As the example of the Bell System electronics engineer illustrated, the detail or length of documents can lead some users to skip ahead when material seems familiar or redundant. Many current documents are so thick they are as forbidding in appearance as in use. Consequently, users sometimes take the attitude Patricia Wright attributed to British housewives dealing with new appliances: When all else fails, read the manual.

Can a usability tester find a truly representative sample to cover both the user who knows too much and the user who knows too little? Can the same document satisfy the needs of both? One attempt to handle this dilemma

is apparently the two-part document. An introductory section lets the impatient user get right to application with a minimum of instruction, and the insecure user choose the second section for more complete instruction and for the rationale behind the program. This can work well if users sort themselves out properly.

Documents are not always clearly step-wise in nature, of course. Documentors can then face the problem of readers who do not understand that they do not understand. How can usability observers reliably discover where the problems occur if readers cannot? Some promising experimental approaches are described below, but they, like protocol analysis, are time-consuming and labor-intensive. Do documentors have the time? Can usability testers routinely apply them?

The problem of readers failing to understand when they do not understand seems to begin early in the process of learning to read. Studies show that children rather quickly develop the notion that something in print must be correct; if text is not clear, they themselves must be to blame. Children can even pass over obvious errors deliberately placed in text (inappropriate terms, misspellings, etc.). When asked about them afterward, children often report that they themselves must have messed up somewhere. This observation raises some questions. Are the respondents in usability testing always willing to acknowledge errors? Can they always admit problems in the text when their inability to understand seems to be simple-minded? Can they be counted upon to blame the text rather than themselves? Or might some respondents be overly critical? Are unnecessary details added to documents to satisfy them?

Mention of text raises other problems with computer documents. One is discrepancy between what a manual and a help screen say. The two may be written from different points of view, especially where different documentors are involved. A more basic discrepancy appears in manuals written on the assumption that readers know something rather than that they are trying to learn something. Procedures can then have missing steps. If hardware is the topic, difficulty may arise when a part included in the package is not addressed in the document. Or a part may be defective, but the user may not have sufficient background to consider this a possibility.

Another problem is more common: lack of hard copy. Many documents nowadays exist only on computer screens; no manuals are provided. Users can print hard copy from these sources if they wish, but are often reluctant to do so for one reason or another. Besides, as Zibell points out, readers are more likely to "trust paper and the information that comes on it." They may become irritated at the lack of a manual for reference, actually deciding against using a program that has none. Perhaps a new generation that has grown up with documents only on screens will not feel strongly about this lack, but many current users do.

Indexes

Indexing of documents, and particularly computer-based indexes, came in for almost universal criticism among my respondents. Ineffective look-ups can occur on first using and learning programs because the index terminology does not always correspond with the terms in users' minds. One of my respondents reported trying to look up information on "matrix analysis functions." None of these words singly or in combination appeared in the index for his program, and only blind luck led to the term "determinant," which gave him access to the information he wanted. Berkeley's COIK criticism is obviously still appropriate.

Usability testing can forestall many such problems, as Redish suggests when she asks of an index, "Does it have users' words in it?" However, the problem arises in another form after the user has experience with a program but again needs help. At that point, the vagaries of long-term memory enter the picture. A term that led to effective look-up during usability testing is sometimes either forgotten or modified in the user's mind. A futile search may then follow, employing ineffective terms that did not

come up during usability testing.

The index for complex programs must necessarily be quite large, so shortcuts are often taken to keep them from getting too large or redundant. One frustrating result is circular reference, where a look-up leads to a term that itself requires look-up, which leads to still another term that requires look-up,etc. If documentors would follow a rule limiting the number of clicks necessary to get to desired information, it would help. A good rule here is the three-click rule; any further links are likely not to be tolerated happily.

How might a documentor provide the background information that prevents this circularity without introducing too much redundancy? Perhaps more use can be made of the Open University's approach: provide key terms the reader needs to know before proceeding. Such terms and their definitions might be placed on the borders of manual pages or, where a colleague and I did in a book on statistics, on fold-out pages. The user can, with this information, decide whether to proceed.

Berkeley's COIK criticism fits another complaint among my respondents: computer jargon. This can take the form of noun strings, such as "Expense Reports Create Screen." This cryptic message can baffle and frustrate a user not familiar with document terminology, just as unexplained error messages can. Still another problem is abbreviations, such as GIF, JPEG, etc., that are easily forgotten after introduction yet are sometimes crucial to understanding. Perhaps the most irritating are abbreviations that also appear hard to read, such as WYSIWYG, which can make a reader feel dyslexic. Unfortunately, just as documentors cannot count on the Web command of Hargis that "Reading is fun," they cannot take assurance from the ironic Web joke "Dyslexics have more fnu."

What can users do when they have trouble comprehending documents? Producers of the offending hardware and software can be called, but they are not as generous as they once were with advice. Books that either simplify or extend the information in a document, even those with offensive titles, provide another expensive source of help. Or users may be fortunate enough to have a helpful help desk in their university or commercial organization. Wherever they turn, cries for help are all too common for both beginners and advanced users of complex documents. Research on the process of comprehension is needed, and some promising approaches exist.

Comprehension Research

Study of causal factors in the comprehension process is emphasized in such work as that of Walter Kintsch and his associates (Kintsch, 1974, and Kintsch and Vipond, 1979, for example). Chall (Chall and Dale, 1995) characterizes this work as being at readability level, with its emphasis on the central role of propositions and the use of reinstatement searches. While propositional analysis and the observance of reinstatements are likely to be too time-consuming for regular usability analysts at this time, the underlying notions can provide helpful suggestions.

Susan Kemper (1983, 1993) has also been working at the readability level of text in studying inference load, a promising notion. (Dot-com entrepreneurs already use inferences in "profiling" to predict what individuals may buy from their personal interests and preferences.) Bonnie Meyer and associates (Meyer, 1975; Meyer and Freedle, 1984) have studied the comprehension of larger prose units in terms of causative (e.g, if - then) relationships (see also Calfee and Curley, 1984). Keeping these relationships in mind is especially important when revising documents. Halliday and Hasan (1976) and Bonnie Armbruster (1984) have explored the issue of coherence in texts, which is of special importance in longer documents.

Brief introductions to the research of these investigators can be found in Klare (1985) and Chall (Chall and Dale,1995). As noted, Chall, in her interpretation of the work of Kintsch, relates it to classic readability measurement. She defends the classic approach of readability by quoting Kintsch in regard to his six-factor "formula." He says, "most of the variance is

accounted for by the first two factors—the number of reinstatements, as first discussed, and the traditional word frequency." Kling and Pratt (1977) also report that *number of syllables* in text yields a higher correlation with Kintsch's passage complexity ratings (.70) than *number of propositions* (.60). As Chall states, causal research on comprehension such as that mentioned above should be encouraged. However, she argues that this should not imply that the classic variables in readability formulas, and indeed formulas themselves, should be abandoned where they can be usefully applied.

The Final Word(s)

Documentors need not be concerned for the significance of their efforts these days. The United States economy and position in the world depend upon staying at the front of the computer-based information revolution. Skilled documentation is more important than ever if the next obstacle, the so-called "digital divide," is to be bridged. Many of the potential users of computers require especially readable documents because they have either limited schooling or limited aptitude for schooling. Writing that appears difficult to read can all too easily turn them off, as it already has in many cases. What can be done to help solve this problem?

We can expect some help from the developers of electronic hardware through attempts to make their products easier to use. We can expect some help as more becomes known about learning and the human mind. Basic research on the brain, such as that being funded by the McGoverns at MIT (2000) and elsewhere, can be expected to lead eventually to more efficient learning. But neither the principle of least effort nor the limitations of short-term and long-term memory are likely ever to be repealed.

There is a limit to how much simplifying the hardware makers can do as equipment handles more and more complex tasks. The rest is up to documentors to make the tasks of users easier. Usability analysis provides a practical approach for documentors, but needs more research for maximum effectiveness in document preparation. Research is also needed to further evaluate the relative effectiveness for reader comprehension of such arrangements as bulleted lists, organization, format, coding, and other such techniques, as well as when best to use them.

Studies already make clear that certain reader characteristics relate to comprehension of documents. A high level of reading skill, when present, can provide a high level of reader comprehension. A high level of motivation can be shown to override some comprehension difficulties, and a high degree of background knowledge can make even difficult documents comprehensible. But documentors cannot always count on high levels of reader skill, motivation, and background.

Here, then, is the domain of readability principles and, where they can apply as screening devices, readability formulas. Readable text consistently results in increased reading speed and efficiency and can lead to increased reader acceptance and persistence. Readable text can also increase reader comprehension often enough to make the effort to produce it worthwhile, especially for the less competent reader. Consequently the final word in this document, like the initial word, must be "readable."

References

Anderson, R. C., and Freebody, P. (1981). Vocabulary knowledge. In J. T. Guthrie (Ed.), *Comprehension and teaching: Research reviews* (pp. 77-117). Newark, DE: International Reading Association.

Anderson, R. C., and Freebody, P. (1983). Reading comprehension and the assessment of word knowledge. In B.Hutson (Ed.), *Advances in Reading/Language Research, 2* (pp. 231-256). Greenwich, CT: JAI Press.

Armbruster, B. B. (1984). The problem of inconsiderate text. In G. Duffey (Ed.), *Comprehension instruction* (pp. 202-217). New York: Longman.

Barzun, J. (1980). Quotation in *The editorial eye,* May, 1980.

Benson, R. W. (1984-1985). The end of legalese: The game is over. *Review of Law and Social Change, 13 (3),* 519-573.

Berkeley, E. C. (1949). *Giant brains, or machines that think.* New York: Wiley.

Berkeley, E. C. (1967). Computer-assisted explanation. *Computers and Automation, 16,* 32-37.

Bormuth, J. R. (1966). Readability: A new approach. *Reading Research Quarterly*, *1,* 79-132.

Brown, W. B. (1965). Science textbook selection and the Dale-Chall formula. *School Science and Mathematics, 65,* 164-167.

Calfee, R. C., and Curley, R. (1984). Structures of prose in the content areas. In J. Flood (Ed.), *Understanding reading comprehension* (pp. 161-180). Newark, DE: International Reading Association.

Calvin, W. H., and Bickerton, D. (2000). *Lingua ex machina.* Cambridge, MA: MIT Press.

Caterpillar Fundamental English (1972, 1976). *Pocket reference.* Peoria, IL: Caterpillar Tractor Co.

Chall, J.S., and Dale, E. (1995). *Readability revisited: The new Dale-Chall readability formula.* Cambridge, MA: Brookline Books.

Charrow, R. P., and Charrow, V. R. (1979). Making legal language understandable: A psycholinguistic study of jury instructions. *Columbia Law Review,* 79, 1306-1374.

Coleman, E. B. (1971). Developing a technology of written instruction: Some determiners of the complexity of prose. In E. Z. Rothkopf and P. E. Johnson (Eds.), *Verbal learning research and the technology of written instruction* (pp. 155-204). New York: Teachers College Press, Columbia University.

Current Contents. (1988). *Social and Behavioral Sciences* and *Arts and Humanities, 20 (47).*

Dale, E., and Chall, J. S. (1948). *A formula for predicting readability.* Columbus, OH: Ohio State University Bureau of Educational Research.

Dale, E., and O'Rourke, J. (1976, 1981). *The living word vocabulary.* Chicago: World Book/Childcraft International.

Duffy, T. M. (1985). Readability formulas: What's the use? In T. M. Duffy and R. Waller (Eds.), *Designing usable texts* (pp. 113-143). Orlando, FL: Academic Press.

Duffy, T. M., and Kabance, P. (1982). Testing a readability approach to text revision. *Journal of Educational Psychology, 74,* 733-748.

Entin, E. E., and Klare, G. R. (1985). Relationships of measures of interest, prior knowledge, and readability to comprehension of expository passages. In B. Hutson (Ed.), *Advances in Reading/Language Research, 3* (pp. 9-38). Greenwich, CT: JAI Press.

Ewing, M. J. (1976). *A comparison of the effects of readability and time on learning the contents of a state driver's handbook.* Unpublished doctoral dissertation, Florida State University.

Fass, W., and Schumacher, G. M. (1978). Effects of motivation, subject activity, and readability on the retention of prose materials. *Journal of Educational Psychology*, *70*, 803-808.

Flesch, R. F. (1943). *Marks of readable style: A study in adult education.* New York: Bureau of Publications, Teachers College, Columbia University.

Flesch, R. F. (1948). A new readability yardstick. *Journal of Applied Psychology, 32,* 221-233.

Flower, L. (1985). *Problem solving strategies for writing.* San Diego, CA: Harcourt, Brace, Jovanovich.

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

Flower, L., Hayes, J. R., Carey, L., Schriver, K., and Stratman, J. (1986). Detection, diagnosis, and the strategies of revision. *College Composition and Communication, 37,* 16-55.

Fry, E. B. (1977). Fry's readability graph: Clarifications, validity, and extension to level 17. *Journal of Reading, 21,* 242-252.

Fry, E. B. (1989). Readability formulas: Maligned but valid. *Journal of Reading, 32,* 292-297.

Greene, M. T. (1979). *Effects of readability and directed stopping on the learning and enjoyment of technical materials.* Unpublished doctoral dissertation, University of North Carolina.

Halliday, M. A. K., and Hasan, R. (1976). *Cohesion in English.* New York: Longman.

Holmquist, J. B. (1968). *A determination of whether the Dale-Chall readability formula may be revised to evaluate more validly the readability of high school science materials.* Unpublished doctoral dissertation, Colorado State University.

Kemper, S. (1983). Measuring the inference load of a text. *Journal of Educational Psychology, 75 (3),* 391-401.

Kemper, S. (1993). Enhancing older adults' reading comprehension. *Discourse Processes, 16* (4), 405-428.

Kern, R. (1999). e-mail reply, April 23.

Kincaid, J. P., and Gamble, L. G. (1977). Ease

of comprehension of standard and readable insurance policies as a function of reading ability. *Journal of Reading Behavior, 9,* 95-87.

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

Kintsch, W., and Vipond, D. (1979). Reading comprehension and readability in educational practice and psychological theory. In L. G. Nilsson (Ed.), *Perspectives on memory research* (pp. 329-365). Hillsdale, NJ: Erlbaum.

Klare, G. R. (1963). *The measurement of readability.* Ames, IA: Iowa State University Press.

Klare, G. R. (1974-75). Assessing readability. *Reading Research Quarterly, 10,* 62-102.

Klare, G. R. (1976). A second look at the validity of readability formulas. *Journal of Reading Behavior, 8,* 129-152.

Klare, G. R. (1979). *Readability standards for Army-wide publications* (Evaluation Report 79-1.). Fort Benjamin Harrison, IN: Directorate of Evaluation, U.S. Army Administration Center.

Klare, G. R. (1984). Readability. In P. D.Pearson (Ed.), *Handbook of reading research* (pp. 681-744). New York: Longman.

Klare, G. R. (1985). *How to write readable English.* London: Hutchinson.

Klare, G. R., Mabry, J., and Gustafson, L. M.(1955). The relationship of style difficulty to immediate retention and to acceptability of technical material. *Journal of Educational Psychology, 46,* 287-295.

Klare, G. R., Rowe, P. P., St. John, M. G., and Stolurow, L. M. (1969). Automation of the Flesch "Reading Ease" readability formula, with various options. *Reading Research Quarterly, 4,* 550-559.

Klare, G. R., Sinaiko, H. W., and Stolurow, L. M. (1972). The cloze procedure: A convenient readability test for training materials and translations. *International Review of Applied Psychology, 21,* 77-106.

Klare, G. R., and Smart, K. L. (1973). Analysis of the readability level of selected USAFI instructional materials. *Journal of Educational Research, 67,* 176.

Kling, M., and Pratt, D. F. (1977). Readability: A re-examination and new directions for further research. Paper presented at the annual meeting of the National Reading Conference, New Orleans, LA.

McLaughlin, G. H. (1966). Comparing styles of presenting technical information. *Ergonomics, 9,* 257-259.

McLaughlin, G. H. (1969). SMOG grading: A new readability formula. *Journal of Reading, 12,* 639-646.

Meyer, B. J. F. (1975). Identification of the structure of prose and its implications for the study of reading and memory. *Journal of Reading Behavior, 7,* 7-47.

Meyer, B. J. F., and Freedle, R. (1984). Effects of discourse type on recall. *American Educational Research Journal, 21,* 221-243.

MIT receives pledge of $350 million to create institute of brain research. (2000). *Chronicle of Higher Education,* March 10, A-40.

Moore, A. C. (1935). Recoiling from reading: A consideration of the Thorndike Library. *Library Journal, 60,* 419-422.

Murphy, D. R. (1947a). Test proves short words and sentences get best readership. *Printer's Ink, 218,* 61-64.

Murphy, D. R. (1947b). How plain talk increases readership 45 percent to 66 per cent. *Printer's Ink, 220,* 35-37.

Nelson, M. J., and Denny, E. C. (1960). *The Nelson-Denny Reading Test.* Boston: Houghton Mifflin.

Readability Master 2000. (1997). *User's Guide*. Cambridge, MA: Brookline Books.

Redish, J. C., and Selzer, J (1985). The place of readability formulas in technical communication. *Technical Communication, 32 (4),* 46-52.

Richards, L. G., and Heller, F. P. (1976). Recognition thresholds as a function of word length. *American Journal of Psychology, 89,* 455-466.

Rodale, J. I. (1978, 1986). *The synonym finder.* New York: Warner Books.

Schriver, K. A. (1997). *Dynamics in document design.* New York: Wiley.

Schumacher, G. M., Klare, G. R., Cronin, F. C., and Moses, J. R. (1984). Cognitive activities of beginning and advanced college writers: A pausal analysis. *Research in the Teaching of English, 18,* 169-187.

Schumacher, G. M., Scott, B. T., Klare, G. R., Cronin, F. C., and Lambert, D. A. (1989). Cognitive processes in journalistic genres. *Written Communication, 6 (3),* 390-407.

Sinaiko, H. W., and Klare, G. R. (1971). *Further experiments in language translation: Read-*

ability of computer translations. Paper P-761. Arlington, VA: Institute for Defense Analyses, Science and Technology Division.

Spache, G. (1974). A new readability formula for primary-grade reading materials. *Elementary School Journal, 53,* 410-413.

Stevens, S. S., and Stone, G. (1947). Psychological writing: easy and hard. *American Psychologist, 2,* 230-235.

Swanson, C. C. (1979). *Readability and top-level structure: Effects on reading comprehension.* Unpublished doctoral dissertation, Arizona State University.

Tobias, S. (1994). Interest, prior knowledge, and learning. *Review of Educational Research, 64 (1),* 37-54.

Waples, D., and Tyler, R. W. (1931). *What people want to read about.* Chicago: University of Chicago Press.

Zipf, G. K. (1949). *Human behavior and the principle of least effort.* Cambridge, MA: Addison-Wesley.