Revue québécoise de linguistique

Varying Approaches to Readability Measurement

Jeanne S. Chall

Volume 25, Number 1, 1996

Lisibilité et intelligibilité

URI: https://id.erudit.org/iderudit/603125ar
DOI: https://doi.org/10.7202/603125ar

See table of contents

Publisher(s)

Université du Québec à Montréal

ISSN

0710-0167 (print)
1705-4591 (digital)

Explore this journal

Cite this article

Chall, J. S. (1996). Varying Approaches to Readability Measurement. *Revue québécoise de linguistique*, *25*(1), 23–40. https://doi.org/10.7202/603125ar

Article abstract

The article discusses the three approaches to readability measurement that have been developed from the early 1900s to the présent—classic readability, cognitive-structural readability, and judgment-qualitative approaches. The classic approaches to readability are the most widely used. They use similar text features to predict readability—some aspects of word difficulty and some measure of sentence complexity. The cognitive-structural approaches are concerned more with the structure of a text and its meaning. The judment-qualitative approaches do not rely on specific features but on a qualitative judgement of overall difficulty.

Each of these approaches is further treated in terms of its underlying theories, the text features and characteristics mesured, its reliability and validity and its practical uses.

Revue québécoise de linguistique, vol. 25, n° 1, 1996, © RQL (UQAM), Montréal

VARYING APPROACHES TO READABILITY MEASUREMENT*

Jeanne S. Chall
Harvard University

1. Introduction

THE ART AND SCIENCE of estimating the comprehension difficulty of spoken and written language has had a long history. It is said that the ancient Greeks looked to experienced orators to predict whether given texts could be understood by different audiences, cf. Chall (1988). Hebrew scholars were known to develop frequency counts of the words in various religious texts, cf. Lorge (1944b).

Historically, judgment preceded objective measurement and until the early 1900s was the most prevalent method used to estimate comprehension difficulty. The wide use of objective measurement began in the United States around the 1920s and it has been the predominant approach since then, although there has been a growing trend recently to return to judgment and qualitative techniques.

In this paper I will be concerned with three approaches to the measurement of text difficulty — two kinds of objective procedures, the classic readability formulas and the more recent cognitive-structural approaches, and judgment and qualitative procedures. I will present the theories underlying each, their features, some evidence of their validity and their practical uses. It is hoped that this review will help researchers and users to gain a greater understanding of the various approaches. The three approaches will be presented in the order of their current use — the classic approach, the cognitive-structural, and the holistic-judgment approach.

---

* This article is a shorter version of a more extensive analysis of readability appearing in *Readability Revisited and the New Dale-Chall Formula* by Jeanne Chall and Edgar Dale published by Brookline Books, Cambridge, MA 1995.

2. Classic Readability

Classic readability dates back to the 1920s. Several comprehensive reviews of classic readability are available: Chall (1958, 1984), and Klare (1963) reviewed the early research and applications. Klare further updated the research from the 1960s to the 1970s. Additional updates are found in Chall (1979, 1981, 1984, 1988), Chall & Conard (1991), Klare (1984,1988), and Chall & Dale (1995).

The classic approach to readability is still the most widely used for predicting comprehension difficulty of materials at the elementary, high school, college and adult levels. There are over 50 classic procedures for predicting text difficulty, usually referred to as readability formulas. But only about 5 have been widely used — Lorge (1944a), Flesch (1948), Dale & Chall (1948), Bormuth (1969), and Fry (1977).

Essentially, these classic readability measures use similar factors to predict comprehension difficulty — some aspect of word difficulty measured either as word familiarity, word frequency, abstract versus concrete words, or word length — number of syllables, number of letters, or affixes, etc. — and some measure of sentence complexity, measured either by average sentence length, or by complex versus simple sentences.

These two factors together — words and sentences — estimate the difficulty of texts when compared to independent measures of text difficulty such as scores on reading comprehension tests, cloze tests, rate of reading, oral reading errors, written summaries, and expert or reader judgments of difficulty. The validity coefficients of these measures have been quite high — from about .6 to .9. Generally, the later the publication, the higher the prediction.

Early on, some readability researchers suggested that conceptual difficulties and idea density were better predictors than semantic and syntactic factors. Some even suggested that conceptual aspects of text were more fundamental than word and sentence measures. But the conceptual measures they proposed were difficult to measure objectively. Also, since the conceptual factors were highly related to the word and sentence measures, they were seldom used.

For example, Ojemann (1934) noted that harder texts contained more abstract words and incoherent expressions. He emphasized, as did the cognitive psychologists in the 1970s and 1980s (see next section), that difficulty comes from ideas rather than from words and sentences. Difficult passages contain hard words because the ideas they express are difficult and abstract. Easy passages contain familiar words because they deal with familiar, concrete ideas.

Ojemann also cautioned that readability factors not be viewed mechanically — cautions expressed again and again in both the classic readability literature and more recently, in the readability literature of cognitive psychologists. Thus, although cognitive features were not incorporated in most of the classic readability formulas, they became part of the broader knowledge of estimating comprehension difficulty.

An early scheme that substituted conceptual for word and sentence factors was developed by Morriss and Holversen in 1938. They classified content words into those that are simple word labels, words that signify concrete ideas, and those that are abstract. A comparison of these ratings with the classic Lorge readability formula resulted in a substantial positive correlation between the two, see Chall (1958).

Still another attempt to "free up" readability measurement from word and sentence complexity was Taylor's (1953) cloze procedure. Taylor proposed that the ability to fill in the correct deleted words in a text was a better predictor of difficulty than the classic readability formulas. For example, his cloze procedure proved to be a better predictor of reading difficulty for a selection from a Gertrude Stein novel because her unusual use of simple words masked the difficulty of her ideas.

The cloze procedure has had a long and fruitful history. But it did not replace the classic readability formulas since it required a panel of readers to take the cloze tests to test readability, while the classic readability formulas required only the counting of words and sentence factors. But the cloze procedure has been extremely useful in the construction of criterion passages — passages of increasing difficulty used to develop readability formulas.

Classic readability formulas have been found to be reliable and valid measures for estimating difficulty — the later ones tend to be more valid than the earlier ones. For example, in the 1958 review by Chall, the validity of most classic formulas ranged from .6 to .7. By the 1990s, the validity coefficients were close to .9, see Hayes, Wolfer & Wolfe (1993), Stenner, Horabin, Smith & Smith (1988b), and Chall & Dale (1995).

Most developers of readability formulas have cautioned against the mechanical use of readability measures as prescriptions for writing and editing. Most have recommended instead that broader aspects of readability be considered as well as word and sentence factors in preparing readable texts. Indeed, guidelines for writing such texts were written by several developers of classic readability formulas, see, for example, Flesch (1949, 1974), Dale & Hager (1950), and Gunning (1952, 1968).

Most researchers on readability agreed that texts could not be simplified or made more difficult merely by changing the words and sentences. This was expressed eloquently in 1937 by Ernest Horn who noted that substituting easier for harder words may not cut down on comprehension difficulty. Indeed, using more high frequency words may even cause greater difficulty.

> *«There is real danger that the mechanical and uncritical use of data on vocabulary will not only effect adversely the production, selection, and use of books but will result in absurdities that will throw research in this field into disrepute.»* Horn, 1937, p. 162

See, for example, Davison & Kantor (1982) for a more recent expression of the same concern that adaptations and revisions of texts by mechanically lowering readability scores can result in less readable text.

Does this mean, then, that classic readability factors have no value in writing readable texts? Not completely. A review of the relevant research from 1920 to 1958 by Chall concluded that benefits in terms of increased comprehension and interest have been demonstrated by simplifying vocabulary and sentence structure. But, «such benefits were found only where gross changes were made or when other, more subtle factors such as organization and directness of approach were also changed», cf. Chall (1958, p. 166).

Klare (1984) came to essentially the same conclusion. He cites, as reasons for the weakness of classic readability in guiding writing, the fact that word and sentence variables are not the only contributors to readability. «Among the more obvious missing candidates are organization, format, and illustrations, (verbal and pictorial)» (p. 717), see also Pearson (1974-1975).

More recently the classic readability formulas have been incorporated into computer programs for writers. One such early program, *The Writer's Workbench*, was developed by Frase and his associates at Bell Laboratories, cf. Frase (1980), MacDonald, Frase, Gingrich & Keenan (1982). Frase provided information on the following aspects of writing, most of which come from classic readability formulas: several readability indexes (grade levels), information on the average lengths of words and sentences, distribution of sentence length, grammatical types of sentences used (simple and complex), percentage of verbs in the passive voice, etc.

Another computer program, cf. Kincaid, Aagard & O'Hara (1980), Kincaid, Aagard, O'Hara & Cottrell (1981), for writers, provides a readability formula score (grade level and related statistics), flags and lists uncommon words, flags

long sentences, and offers the writer simpler and more common options in the text itself alongside the uncommon words and phrases.

3. The New Cognitive-Structural Readability

About the middle of the 1970s, discontent with classic readability seemed to grow quite strong, although the readability formulas continued to be used widely. The major discontent with the classic readability formulas expressed by the critics was that they measured "mere surface factors," not real sources of difficulty. Some critics acknowledged that the classic formulas had practical validity and value, but were essentially a-theoretical. What was needed instead, it was claimed, were schemes based on cognitive theory that would explain text difficulty, and could be used more effectively for estimating difficulty and for writing and editing readable text, cf. Kintsch & Vipond (1979).

The work of Kintsch and his associates is used here to illustrate the cognitive approach to readability and that of Meyer to illustrate the importance of organization.

4. Kintsch on Classic and Cognitive Readability

Kintsch presented his main ideas on classic and cognitive readability in a series of articles written between 1977 and 1981. In these he notes that his intentions are to add to the classic formulas, not to supplant them, cf. Kintsch & Vipond (1977[1], p. 11).

In the 1977 article, he is quite critical of classic readability — that it is a-theoretical, is based only on text factors, and overlooks the interactive aspects of text difficulty with reader characteristics. He also questions the value of the classic readability factors — word difficulty and sentence complexity — for determining text difficulty, and especially for writing and editing texts at specified difficulty levels.

I am tempted here to respond to the criticisms, but hesitate lest it be viewed as questioning the cognitive-structural approach. As noted earlier, cognitive and structural factors were found important from the beginnings of classic readability. But new measures do not necessarily invalidate old ones. It is in this spirit that I comment on Kintsch's position on classic readability.

[1] We cite from the 1977 manuscript of the paper presented by the authors. It can also be found in published form (1979).

If his claim is that classic readability is not grounded in modern cognitive psychology, it is a valid point. Classic readability came into being about 50 years before modern cognitive psychology. But classic readability has perhaps a longer theoretical and research base — that of the development of language and of reading comprehension, see, for example, R.L. Thorndike (1973-74).

That classic readability cannot be turned directly into rules for writing has long been known by classic readability researchers (see above). But the recent work of Kintsch and his associates suggests that cognitive-based readability is also subject to similar weaknesses when turned into rules for writing, cf. Kintsch, Britton, Fletcher, Kintsch, Mannes & Mitchell (in press).

The claim that classic readability is not interactive with reader characteristics is puzzling. The traditional readability measures have used reading levels or qualitative descriptions of readers which help writers understand their potential readers, see Chall & Dale (1995), Chall & al. (1996), Carver (1990).

That cloze tests and even multiple choice comprehension tests and reader judgments are not valid as criteria of difficulty is also puzzling. There is considerable evidence that quite similar results are obtained from multiple-choice tests of reading comprehension, cloze tests, judgments of difficulty, and the summaries and reading rate which Kintsch preferred in his early articles. In his most recent article, in fact, Kintsch reports that substantial correlations were found on free recall short answer tests (immediate and delayed) and learning from text, cf. Kintsch & al. (1993). See also Chall (1958), Chall & Conard (1991, and Stenner, Horabin, Smith & Smith (1988a).

Overall, the two broad and important factors affecting readability proposed by Kintsch and Vipond (1977) are the propositional density of a text and the number of new concepts per proposition. More, specifically, the factors included in their analysis were: density of propositions, and the number of different arguments, coherent parts, inferences required to connect a text base, long-term memory searches and reinstatements of propositions into short term memory, and reorganizations required to arrive at the best organized text base.

Kintsch and Vipond present the results of analyses using cognitive measures as well as those from classic readability. They report that the classic factors were indeed good predictors of readability. The six predictor variables — the number of reinstatement searches made in processing the paragraph, the average word frequency (a common classic readability factor), the proposition density, the number of inferences, the number of processing cycles, and the number of different arguments in the proposition list correlated "a proud .97" with a test of immediate recall. «Most of the variance is accounted for by the

first two factors — the number of reinstatements... and the traditional word frequency... That word frequency and sentence length are related to reading difficulty is not news», cf. Kintsch & Vipond (1979, p. 10). Thus, it appears that a combination of a new cognitive factor (the number of reinstatements) and a classic factor (word frequency) were the strongest in predicting text difficulty.

The 1980 article by Miller and Kintsch contains an extended and more refined analysis of Kintsch's cognitive approach to readability. It reiterates the notion that readability is an interaction between texts and readers and that difficulty in reading stems from locating and maintaining relationships between ideas. This is best shown by the increased time needed to read the material, by the amount recalled of the material read, and the time per unit of information recalled.

They also present the results of their analysis of 20 passages from the Reader's Digest for the number of inferences, arguments cycles, words per proposition, short-term memory stretches, input and buffer size along with measures of word frequency, sentence length, Flesch readability scores, and a rating of subjective readability. The multiple correlations of these measures were .83 for reading time, and .85 for recall.

The predictive power of the classic readability variables was again found to be quite high in this analysis. «The direction of the correlations... confirm the expectation that low reading times and high recall should be characteristic of texts with common words and short sentences», cf. Miller & Kintsch (1980, p. 347).

In spite of the high correlations the authors beg off producing a new readability formula. Instead, they say that the high correlations «should be considered as indications that certain model-dependent predictors, primarily reinstatement and inferences, are indeed important determiners of readability» (p. 348).

The Kintsch articles published from 1977 to 1981 present an interesting development of his theories and views. The first article, published in 1977, appears to be quite uncompromising in its criticism of classic readability.

By the second and third articles, their position seems to soften. In fact, Kintsch and his associates used classic readability factors along with their cognitive factors and found both to be quite valid in relation to their cognitive factors. Thus, in the 1980 article we read that «the readability of a text is determined by the ways that certain text properties — primarily the arrangement of the propositions in the text base, but also the word frequency and sentence length — interact with the reader's processing strategies and resources», cf. Miller & Kintsch, (1980, p. 348).

In the Kintsch and Miller article of 1981, we have the most comprehensive view of their cognitive model of readability. Although classic readability is still treated with some reservation, it is presented in greater detail, thus making it possible for the reader to make his own interpretations. They also note that the classic readability formulas do work and why they work:

> *«The reason that readability formulas have worked at all is that the factors that make up these formulas are indeed correlated with the conceptual properties of texts: long sentences generally correspond to complex syntactic structures, infrequent words generally refer to complex concepts, and hard texts will generally lead to harder questions about their content.»*
>
> Kintsch & Miller, 1981, p. 222

A later work, cf. Kintsch & al. (in press) presents some important insights into the different effects of text organization on readers of different reading ability. They found, for example, that 6th graders who read materials at a 6th grade level performed better on passages with better micro and macro structures. Tenth graders, as expected, outperformed the 6th graders on the same selections, but were not as affected by the structure of the texts. They read the originals as well as the revised versions. Most surprising to the researchers was that college students wrote better summaries on the poorly written text.

> *«This rather surprising result has some important educational implications which, however, pose a difficult challenge when it comes to designing appropriate kinds of instruction. On the one hand, less skilled readers and those with little background knowledge in a domain need maximum support. One way to provide this would be to construct very explicit, coherent texts which reduce the amount of gap-filling inferences needed to form a coherent representation of the content.. On the other hand, readers with adequate literacy skills who are moderately familiar with the domain might benefit a great deal from having to work harder to get the meaning. By breaking down easy, automatic processing and increasing the amount of active, constructive effort needed to understand a text, learners are forced to engage in more problem-solving activities, these, in turn, may help them achieve a deeper understanding of what they are reading... Students may be better served by an instructional approach that is sensitive to the interactions of text quality and individual differences in readers, but one that helps students do their own thinking, rather than doing it for them.»*
>
> Kintsch & al., in press

We see from the above that optimal challenge is central to both readabilities — the classic and the cognitive. While some early critics of classic readability suggested that this issue would be solved by cognitive readability, we find that

it exists for the new measures as for the old. Compare Chall, Conard & Harris (1977) and Chall & Conard (1991).

To sum up the work of Kintsch and his associates on cognitive readability: Their major objective was to develop a theory of text comprehension — and with it, to explain as fully as possible how readability and comprehension work. They took the position that this is more essential than developing a better technique for readability measurement.

Their explanations of readability are compelling and are generally supported by their theory and their research findings. But one is also struck by the complexity of their scheme for practical use. Although their early articles are strongly critical of classic readability, they do use the main classic readability factors as well as their cognitive measures.

5. Readability of Larger Units of Text: The work of Bonnie Meyer

During the middle and late 1970s, some researchers began serious study of the effects of larger units of text. The researcher who contributed most to this area, and who perhaps had the strongest influence on the work of others, is Bonnie Meyer.

In an article published in 1982, Meyer reports that a communication is vastly more efficient (it saves effort) and is more effective (it gets better results) if it follows a topical plan instead of being a miscellaneous sequence of sentences or paragraphs:

> *«That is, people remember more and read faster information which is logically organized with a topical plan than they do when the same information is presented in a disorganized, random fashion... Thus the plan of a discourse can be considered apart from content, and deserves separate consideration from researchers, as from those who are planning a composition.»*
>
> Meyer, 1982, p. 38

It is important to note that Meyer speaks of structure as being separate from, not in place of, other ways of measuring content — e.g. the microanalysis of Kintsch and his associates. Her work suggests that the presence of a visible plan for presenting content plays a key role in assessing the comprehension difficulty of text and that a text displays a hierarchy of content so that some facts (statements, etc.) are superior or subordinate to others. Such a hierarchy must be based on a plan, and readers who use a plan different from the author's may be at a disadvantage.

Drawing upon linguistics and rhetoric, she gathered empirical evidence about writing plans: e.g., antecedent/consequence; comparison/contrast; description; response; and time order. The *antecedent/consequent* plan is devoted to presenting causal relationships like "if/then." The *comparison* plan presents two opposing viewpoints which give equal weight to both sides, or the *adversative* which clearly favors one side over the other. The *description* plan develops a topic by describing its component parts, for instance, by presenting attributes, specifications or settings. The *response* plan contains some kind of statement followed by a response, such as remark and reply, question and answer and problem and solution. The *time order* plan relates events or ideas according to chronology.

Political speeches, she notes, are often of the comparison type, particularly the adversative subtype. Newspaper articles are often of the descriptive type and scientific papers often use the response type (raising a question or problem and then seeking to give an answer or solution.) History texts frequently use the time order plan.

Meyer found that better readers tend to use the same plan as the authors of the material they are reading. Further, they remember more of the text on immediate and delayed recall.

Some of her ongoing studies suggest that the descriptive plan is the least effective when people read or listen to text for the purpose of remembering it. The comparison and the antecedent/consequent plans are better than the descriptive for identical content. There were also big differences in the kinds of information remembered based on the organizational plan of the writing.

Meyer also distinguishes two types of highlighting of information: *subordination* and *signaling*. For example, in subordination, a main idea can be supported with reasons for the reader to believe it. Or a writer may describe an object in ever greater detail. Signaling is done with explicit markers, such as "on the one hand/or the other hand", "three things must be stressed here", and other expressions indicating how the content is organized.

In Bonnie Meyer's more recent work (1993) she found significant correlations between readability scores (based on her discourse processing analysis) and test item scores on the ETS Basic Skills Test among older adults.

> *«Theoretically, it suggests that prior research on the dimensions of text readability are salient for predicting actual comprehension in everyday task materials for older adults... Practically, the text processing model outlined in the present paper presents a useful algorithm for the selection of test items.»* Meyer, 1993, p. 25-26

These findings, Meyer noted, are quite significant. The ability of the discourse factors to predict comprehension of everyday tasks is further confirmation for their broad theoretical and practical value in predicting readability. It also suggests that readability analysis may be used as a measure to predict test item difficulty. At present, most item difficulty is obtained only by pilot testing. See in this connection, Chall & Dale (1995).

We see in Meyer's work a synthesis of what has been sensed about clear and effective writing for thousands of years. Her studies of text comprehension have been concerned mainly with the broader organizational aspects of text, but her more recent papers have also incorporated the findings from classic readability.

6. The Use of Judgment to Assess Readability

In reality, the use of judgment for estimating readability is probably the oldest approach. Yet one finds little reference to its use except for the general acknowledgment of the excellent grading of the McGuffey readers. Reports on the success of McGuffey note that he graded his books according to his best judgment and that of experienced teachers.

Judgment has also been used in the development of several of the classic readability formulas to determine the independent difficulty of their criterion passages, and also to help validate the levels of difficulty derived from the statistical analyses, cf. Chall (1958).

That judgment can be valid has been known for quite some time in psychological and educational research, see E.L. Thorndike's writing scale (1912). More recently, Shapiro (1967) found that adults estimates of the frequency of words correlated .9 with the frequencies found on the Thorndike-Lorge list (1944). Porter and Popp (1975) found a correlation of .8 between judgments of difficulty of children's books and the difficulty of those books as measured by Cloze scores and oral reading errors. Earlier, Chall (1958) found a .8 correlation between judges' ratings of the difficulty of passages and the Dale-Chall readability levels.

Qualitative or holistic assessment has become of even greater interest in the past decade. The use of portfolios to assess student writing development relies strongly on teachers' judgments. The current use of trade books for teaching reading has produced several qualitative schemes for assessing text difficulty, see Clay (1991).

The Reading Recovery Program at Ohio State University has developed a list of titles for the early grades based on recommendations from New Zealand's Reading Recovery program and the results of field tests with teachers and children in the United States. As new books become available they are "leveled" against this basic list and trial use by children and teachers.

> *«The Ohio State Reading Recovery Booklist is organized along a continuum of 20 levels, ranging in complexity from simple caption texts to stories similar to those found in first grade readers. Levels are only approximate indicators of text difficulty and are intended to serve as a guide for the Reading Recovery teacher in the selection of the new book. Text features that influence the choice of a level for a particular book include: content in relation to children's personal experiences and interests; repetition of language patterns; vocabulary; illustration support for the meaning of the text; narrative style; and the size and placement of the print.»*
>
> Reading Recovery Booklist, August 1990, p. 2

Peterson's (1991) scheme for leveling books for the early grades is also qualitative and based on the work done in New Zealand, see Clay (1991). It takes account of content, language, and physical features of the text, and describes these and other characteristics at five broad levels.

Earlier procedures for judging difficulty were developed by Singer and by Carver in the 1970s. In 1975 Singer published his SEER Technique which used passages at increasing levels of difficulty against which various texts could be compared, using judgment. In 1975-76 Carver published a similar scheme for measuring comprehension difficulty, using a scale independently tested for difficulty.

A similar scheme that uses judgment was developed by Chall, Bissex, Conard & Harris-Sharples (1996). They use six scales — 2 each for fiction (literature and popular fiction), social studies (narrative and expository), and science (physical and biological). Readability levels for books are obtained by comparing samples of the text to appropriate passages in a scale.

The greatest value of these schemes is in the short time taken to obtain readability estimates. The classic measures, as a group, take more time than judgment, but considerably less time than the cognitive measures — which are the most time consuming.

Validation of the judgment schemes is not as extensive as for the classic and cognitive-structural procedures. The qualitative-judgment schemes that use scaled examples do present some evidence of validity which is quite high. The "leveling" procedures do not seem to rely on statistical analyses for presenting evidence of validity.

If one accepts the judgment schemes as valid, one may ask why? Perhaps it is the extensive linguistic experience of those who use these schemes. Indeed, most teachers and writers, by virtue of their long experience with language, become ever more sensitive to its uses, and its relative difficulty.

Qualitative assessments can also be more sensitive to the overall difficulty of a text than measures based on text features alone. There are always limits to the text features that can be analyzed. Whether an assessment focuses on word or sentence features or on organization and cognitive features, it leaves out some important variables. Qualitative, holistic assessment can be more sensitive to the great variety of text variables including vocabulary, syntax, conceptual load, text structure, and cohesion that differentiate the levels of texts, see Chall & Dale (1995).

The holistic assessment schemes, because they are based on a total reaction to the text, can take account of these and other factors known to be associated with difficulty. Further, the scales, particularly those for different content, make more transparent what actually happens as writing becomes more mature, more complex and more difficult to read and understand. Thus use of this procedure educates the user while assessing the text. Those using the scales are likely to gain a fuller understanding of the nature of that text and of its comprehension demands on a reader than they would from using other measures of difficulty.

A qualitative-judgment scheme also makes use of the intuitive knowledge of those for whom it is intended — teachers, writers, editors, and others who select, write, or assess materials for readers of given ability and interests. It does not require counting text features. Instead, it matches samples of texts to exemplars scaled on the basis of teacher and student judgments, cloze comprehension tests, and other standard quantitative measures.

7. The Three Approaches to Readability: A Comparison

The greatest difference among the three procedures is in their primary concern for either practice or theory. The cognitive-structural approach focuses most heavily on theory, specifically on cognitive and linguistic theory. Classic readability and the qualitative-judgment schemes are concerned more with practical use. Although many of the cognitive readability researchers have claimed that all but the cognitive readability schemes are a-theoretical, an analysis of the classic and judgment schemes finds that they too have a theoretical base — in language development and in reading comprehension theory. Indeed, when estimates derived from cognitive measures are compared

to those from classic readability, the scores are quite similar. Comparisons of the readability levels derived from qualitative judgments also tend to be correlated positively with those from classic readability formulas.

There have been differences in the factors that each has used to measure text difficulty. Classic readability has tended to measure aspects of word difficulty and sentence complexity. Cognitive-structural readability has measured ideas and organization — their difficulty and the relationships between them. The qualitative-judgment schemes have relied mainly on total impressions although they tend to rely also on descriptions of text at various levels.

There are differences in the ease of using the three readability measures. Whenever the length of time required to analyze a selection for difficulty is mentioned in the literature, there is general agreement that the judgment and classic measures take less time than the cognitive-structural measures.

The classic and cognitive approaches also seem to vary with regard to levels they are most appropriate for. Thus, the cognitive-structural based readability measures seem to discriminate better among materials at higher levels of difficulty. Classic readability measures have been interested in a fuller range of difficulty, leaning perhaps more towards the lower rather than upper levels. For the lower reading levels, words and sentences seem to be of greater importance than the complexity and density of ideas. Evidence on qualitative judgment schemes is not yet available although it would probably be more easily used for materials at lower levels of difficulty.

The validity evidence indicates quite high validity for the classic and cognitive approaches — .92 for the classic new Dale-Chall formula and .93 for Lexile, see Stenner & al. (1988b). The predictive validity of one of the Kintsch cognitive "formulas" is slightly higher — .97 for 6 factors, including the traditional word frequency. The new Dale-Chall and the Lexile use only 2 factors each — words and sentences.

8. Future Trends

An analysis of the three approaches to readability suggests that a synthesis of the strong features of each can do much to improve our theoretical knowledge and practical uses of readability. It is unfortunate that there have been so many negative positions taken towards approaches that are not one's own. And yet combining the cognitive and the classic, as done by Kintsch, results in greater knowledge and in a more useful instrument for estimating difficulty. It would seem that the classic formulas would benefit from including some of the

cognitive-structural factors, and the "leveling" schemes would benefit from the use of some of the well known simple objective procedures to give teachers more confidence in their use.

It is hoped that a synthesis can also bring about a greater simplification in the various readability schemes. At present, the cognitive measures appear to be very complex, time consuming, and expensive for practical use. It would be important to find simpler ways of measuring the cognitive factors — perhaps by using some of the measures found useful in the classic and in the various qualitative-judgment schemes.

For the above reasons, I suggest that future research on text difficulty attempts to combine the useful aspects of the classic, cognitive-organizational, and qualitative-judgment readability. Indeed, we have attempted to do so in the new Dale-Chall readability formula (1995). It uses the classic word familiarity and sentence length factors and supplements these with cognitive-organizational and judgment factors.

References

BORMUTH, J.R. (1969) *Development of Readability Analyses*, Final Report, Project n° 7-0052, Washington, DC, Department of Health, Education and Welfare, ERIC Document Reproduction Service, n° 054 233.

CARVER, R.P. (1975-76) «Measuring prose difficulty using the reading scale», *Research Quarterly*, vol. 11, n° 4, p. 660-685.

CARVER, R.P. (1990) «Predicting accuracy of comprehension from the relative difficulty of the material», *Learning and Individual Differences*, vol. 2, p. 405-422.

CHALL, J.S. (1958) *Readability: An Appraisal of Research and Application*, Bureau of Educational Research Monographs, n° 34, Columbus, OH, Ohio State University Press, reprinted (1974) by Bowker Publishing, Epping, Essex, England.

CHALL, J.S. (1979) «Readability: In search of improvement», *Publishers Weekly*, vol. 216, n°18, p. 40-41.

CHALL, J.S. (1981) «Readability: Conceptions and misconceptions», *SLATE Newsletter*, Urbana, IL, National Council of Teachers of English, p. 1-4.

CHALL, J.S. (1984) «Readability and prose comprehension: Continuities and discontinuities», in J. Flood (ed.), *Understanding Reading Comprehension*, Newark, DE, International Reading Association, p. 233-264.

CHALL, J.S. (1988) «The beginning years», in B.L. Zakaluk & S.J. Samuels (eds), *Readability: Its Past, Present and Future*, Newark, DE, International Reading Association, p. 2-13.

CHALL, J.S., G. BISSEX, S.S. CONARD & S. HARRIS-SHARPLES (1996) *Qualitative Assessment of Text Difficulty: A Practical Guide for Teachers and Writers*, Cambridge, Brockline Books.

CHALL, J.S. & S.S. CONARD (1991) *Should Textbooks Challenge Students. The Case for Easier or Harder Books*, New York, Teachers College Press.

CHALL, J.S., S.S. CONARD & S. HARRIS (1977) *An Analysis of Textbooks in Relation to Declining SAT Scores*, Princeton, NJ, Educational Testing Service and the College Entrance Examination Board.

CHALL, J.S. & E. DALE (1995) *Readability Revisited and the New Dale-Chall Readability Formula*, Cambridge, Brookline Books.

CLAY, M.M. (1991) *Becoming Literate: The Construction of Inner Control*, Portsmith, NH, Heinemann.

DALE, E. & J.S. CHALL (1948) *A Formula for Predicting Readability*, Columbus, OH, Ohio State University, Bureau of Educational Research, reprinted from *Educational Research Bulletin*, n° 27, p. 11-20, 37-54.

DALE, E. & H. HAGER (1950) *Some Suggestions for Writing Health Materials*, New York, National Tuberculosis Association.

DAVISON, A. & R.N. KANTOR (1982) «On the failure of readability formulas to define readable texts: A case study from adoptions», *Reading Research Quarterly*, n° 17, p. 187-209.

FLESCH, R. (1948) «A new readability yardstick», *Journal of Applied Psychology*, n° 32, p. 221-233.

FLESCH, R. (1949 et 1974) *The Art of Readable Writing: With the Flesch Readability Formula*, New York, Harper & Row.

FRASE, L.T. (1980) *Computer Aids for Writing and Text Design*, Symposium presented at the annual meeting of the American Educational Research Association, Boston.

FRY, E.B. (1977) «Fry's Readability Graph: Clarifications, validity, and extension to level 17», *Journal of Reading*, n° 21, p. 242-252. (Hand calculator for the Fry readability scale (extended) available from Jamestown Publishers, P.O. Box 6713, Providence, RI 02940.)

GUNNING, R. (1952 et 1968) *The Technique of Clear Writing*, New York, McGraw-Hill.

HAYES, D., L. WOLFER & M. WOLFE (1993) «Was the decline in SAT scores caused by simplified school exts?», Paper presented at the *Annual Meeting of the American Sociological Association*, Miami, FL.

HORN, E. (1937) *Methods of Instruction in the Social Studies*, New York, Charles Scribners Sons.

KINCAID, J.P., J.A. AAGARD & J.W. O'HARA (1980) *Development and Test of a Computer Readability Editing System*, CRES, TAEG Report n° 83, U.S. Navy Training Analysis and Evaluation Group, Orlando, FL.

KINCAID, J.P., J.A. AAGARD, J.W. O'HARA & L.K. COTTRELL (1981) «Computer readability editing system», *TEEE Transactions on Professional Communications.*

KINTSCH, W. & D. VIPOND (1977) «Reading comprehension and readability in educational practice», Paper presented at the *Conference on Memory*, June 1977, University of Uppsala, (see the published version, 1979).

KINTSCH, W. & E. VIPOND (1979) «Reading comprehension and readability in educational practice and psychological theory», in L.G. Nilsson (ed.), *Perspectives on Memory Research*, Hillsdale, NJ, Erlbaum, p. 329-365.

KINTSCH, W. & J.R. MILLER (1981) «Readability: A view from cognitive psychology», in *Teaching: Research Reviews*, Neward, DE, International Reading Association, p. 220-232.

KINTSCH, W., B.K. BRITTON, C.R. FLETCHER, E. KINTSCH, S.M. MANNES & J.N. MITCHELL (in press) *A comprehension-based approach to learning and understanding.*

KLARE, G.R. (1963) *The Measurement of Readability*, Ames, IA, Iowa State University Press.

KLARE, G.R. (1984) «Readability», in P.D. Pearson (ed.), *Handbook of reading research*, New York, Longman, p. 681-744.

KLARE, G.R. (1988) «The formative years», in B.L. Zakaluk & S.J. Samuels (eds), *Readability: Its Past, Present, and Future*, Newark, DE, International Reading Association, p. 14-34.

LORGE, I. (1944a) «Predicting readability», *Teachers College Record*, n° 45, p. 404-419.

LORGE, I. (1944b) «Word lists as a background for communication», *Teachers College Record*, n° 45, p. 343-352.

MACDONALD, N.H., L.T. FRASE, P.S. GINGRICH & S.A. KEENAN (1982) «The writer's workbench: computer aids for text analysis», *Educational Psychologist*, n° 17, p. 172-179.

MEYER, B.J.F. (1982) «Reading research and the composition teacher: The importance of plans», *College Composition and Communication*, vol. 33, n° 1, p. 37-49.

MEYER, B.J.F. (1993) «Text processing variables predict the readability of every day documents read by older adults», *Reading Research Quarterly*, vol. 28, n° 3, p. 234-249.

MILLER, J.R. & W. KINTSCH (1980) «Readability and recall of short prose passages: A theoretical analysis», *Journal of Experimental Psychology: Human Learning and Memory*, n° 6, p. 335-354.

MORRISS, E.C. & D. HOLVERSEN (1938) «Idea Analysis Technique», unpublished paper, Teachers College, Columbia University, New York, in J.S. Chall (1958), *Readability: An Appraisal of Research and Application*, Columbus, OH, Ohio State University Press.

OJEMANN, R. (1934) «The reading ability of parents and factors associated with the reading difficulty of parent education materials», in G.D. Stoddard (ed.), *Researchers in Parent Education, II*, Iowa city, IA, University of Iowa, p. 19.

PEARSON, P.D. (1974-75) «The effects of grammatical complexity on children's comprehension, recall, and conception of certain semantic relations», *Reading Research Quarterly*, n° 10, p. 155-192.

PETERSON, B. (1991) «Selecting books for beginning readers», in D.E. Deford, C.A. Lyons & G.S. Pinnell (eds), *Bridges to Literacy: Learning from Reading Recovery*, Portsmith, NH, Heinemann, p. 119-138.

PORTER, D. & H.M. POPP (1975) *Measuring the readability of children's trade books,* Report to the Ford Foundation, Reading Recovery Booklist (1990), Columbus Ohio, Ohio State University.

SHAPIRO, B. (1967) *The Subjective Scaling of Relative Word Frequency,* unpublished report to the Ford Foundation.

SINGER, H. (1975) «The SEER Technique: A non-computational procedure for quickly estimating readability levels», *Journal of Reading Behavior,* n° 3, p. 255-267.

STENNER, A.J., I. HORABIN, D.R. SMITH & M. SMITH (1988a) «Most comprehension tests do measure reading comprehension: A response to McLean and Goldstein», *Phi Delta Kappan,* p. 765-767.

STENNER, A.J., I. HORABIN, D.R. SMITH & M. SMITH (1988b) *The lexile framework,* Durham, NC, Metametrics.

TAYLOR, W. (1953) «Cloze procedure: A new tool for measuring readability», *Journalism Quarterly,* n° 33, p. 42-46.

THORNDIKE, E.L. (1912) *Handwriting,* New York, Teachers College Press.

THORNDIKE, E.L. & I. LORGE (1944) *The Teacher's Word Book of 30,000 Words,* New York, Bureau of Publications, Teachers College, Columbia University.

THORNDIKE, R.L. (1973-74) «Reading as reasoning», *Reading Research Quarterly,* n° 9, p. 135-147.