R E P O R T R E S U M E S

ED 011 480 RE 000 027

READABILITY--A NEW APPROACH.

BY- BORMUTH, JOHN R.

PUB DATE 66

EDRS PRICE MF-$0.09 HC-$2.20 55P.

DESCRIPTORS- *READING RESEARCH, *PREDICTIVE MEASUREMENT, *CLOZE PROCEDURE, *READING DIFFICULTY, *LINGUISTIC PATTERNS, GRADE 4, GRADE 5, GRADE 6, GRADE 7, GRADE 8, READING COMPREHENSION, SYNTAX, DATA PROCESSING, READING TESTS, READING ABILITY, TEST VALIDITY, READABILITY, READABILITY FORMULAS, LITERARY STYLE, DALE CHALL LIST OF 3000 FAMILIAR WORDS, STANFORD ACHIEVEMENT TEST, CALIFORNIA READING TEST, WASCO

IN AN EFFORT TO DEVELOP ACCURATE FORMULAS FOR PREDICTING AND CONTROLLING THE READABILITY OF LANGUAGE, FIVE BASIC PROBLEMS WERE STUDIED--(1) THE VARIOUS FEATURES OF WRITING STYLE AS LINEARLY RELATED TO COMPREHENSION DIFFICULTIES, (2) THE CHANGE OF STRENGTHS OF RELATIONSHIPS AS A FUNCTION OF READING ABILITY, (3) READABILITY PREDICTION MADE UPON SMALL LANGUAGE UNITS, (4) READABILITY ACCURACY BASED UPON USE OF LINGUISTIC ANALYSIS TECHNIQUES IN CONSTRUCTION OF FORMULAS, AND (5) A TEST OF USEFULNESS OF CERTAIN LINGUISTIC VARIABLES IN PREDICTING LANGUAGE DIFFICULTY. TWENTY PROSE PASSAGES WERE ANALYZED TO DETERMINE THE DIFFICULTIES OF 5,181 WORDS, 405 INDEPENDENT CLAUSES, AND 365 SENTENCES. SAMPLES WERE CHOSEN FROM FOUR PASSAGES IN EACH OF FIVE SUBJECT AREAS, USING THE DALE-CHALL SCALE FOR GRADES 4 THROUGH 8. FIVE CLOZE TEST FORMS FROM EACH PASSAGE WERE CONSTRUCTED BY DELETING EVERY FIFTH WORD. SUBJECTS, THE ENTIRE ENROLLMENT IN GRADES 4 THROUGH 8 OF ONE SCHOOL DISTRICT IN WASCO, CALIFORNIA, WERE MATCHED FOR FIVE FORM GROUPS. THE TESTING PERIOD LASTED OVER 11 SCHOOL DAYS. READING ABILITY WAS DETERMINED BY RESULTS ON THE STANFORD ACHIEVEMENT TEST--READING, FORM J, AND STUDIED IN RELATION TO THE INTERACTION OF FORM GROUPS ON MEASURES OF WORD, INDEPENDENT CLAUSE, SENTENCE, AND PASSAGE DIFFICULTY. EACH AREA WAS ANALYZED IN DETAIL AND TABULATED FOR STUDY IN LIGHT OF THE LINGUISTIC VARIABLES SET FOR THE STUDY. A BIBLIOGRAPHY OF 31 ITEMS IS ATTACHED. THIS ARTICLE IS PUBLISHED IN THE "READING RESEARCH QUARTERLY," VOLUME 1, SPRING 1966. (MC)

Reading Research Quarterly, Spring 1966, Volume 1, pp. 79-132.

Readability: A new approach

JOHN R. BORMUTH
University of California, Los Angeles

FIVE PROBLEMS basic to the development of precise readability formulas were investigated. Cloze tests were used to determine the comprehension difficulties of 20 passages and of each word, independent clause, and sentence within each passage. A large number of entirely new linguistic variables were derived along with some previously studied variables and some refined versions of previously studied variables. The results were as follows: First, nonlinear correlation techniques will have to be used in some readability formulas. At the word level of analysis, F tests of linearity showed most regressions were curvilinear. The results were inconclusive at the other levels of analysis. Second, readability formulas can predict difficulty as well for subjects at one level of ability as for subjects at the other levels. Third, readability formulas can make usefully valid predictions of the difficulties of individual words, independent clauses, and sentences. Fourth, the validity of readability formulas based entirely on linguistic variables can be greatly improved. A multiple correlation of .934 was found between the linguistic variables and passage difficulty. Fifth, the greatest improvement in readability prediction will result from developing more sophisticated linguistic variables. Several new variables exhibited correlations exceeding .7 with comprehension difficulty. Two correlations reached .8.

La facilité de lire: Une nouvelle approche

CINQ PROBLÈMES fondamentaux du développement de formules précises pour évaluer la facilité de lire fûrent examinés. Des épreuves "Cloze" fûrent employée pour determiner les difficultés de compréhension de 20 passages et de chaque mot, de membres de phrases indépendentes, et des phrases dans chaque passage. Un grand nombre de variables linguistiques entièrement nouvelles fûrent déduites avec quelques variables précèdement étudiées et quelques versions raffinées de variables etudiées précèdement. Les résultats fûrent les suivants: premièrement des techniques de

This study was sponsored by the U. S. Office of Education, Cooperative Research Project No. 2082. The author also wishes to acknowledge the services and assistance rendered by the U.C.L.A. Computer Facility.

corrélation non-linéaire doivent être employées dans quelques formules de la facilité de lire. Au niveau du mot de l'analyse les tests F de linéarité montrant que le plus des rebroussement étaint curvilinéaires. Deuxièmement, les formules de facilité de lire peuvent prédire des difficultés tout aussi bien pour des sujets à un certain niveau de capacité que pour des sujets à d'autres niveaux. Troisièmement, les formules de facilité de lire peuvent donner des prédictions utilement valides des difficultés de mots individuels, de membres de phrases indépendentes et de phrases. Quatrièmement, la validité de formules de facilité de lire basée entièrement sur des variables linguistiques peut être amélioré énormément. Une corrélation multiple de .934 fut trouvée entre les variables linguistiques et les difficultés de certains passages. Cinquièmement, le plus grande perfectionnement pour la prédiction de la facilité d'être lu résultera d'un développement de variables linguistiques plus sophistiquées. Plusieurs nouvelles variables montrent des corrélations plus grandes que .7 avec la difficulté de comprendre. Deux corrélations montrent jusqu'a .8.

Legibilidad: Un nuevo adelanto

SE INVESTIGARON cinco problemas fundamentales sobre el desarrollo de fórmulas precisas de legibilidad. Se utilizaron las pruebas Cloze para determinar las dificultades en comprensión de 20 pasajes y de cada palabra, cláusula independiente, y oración dentro de cada pasaje. Se obtuvo un gran número de variables linguísticas completamente nuevas además de otras variables anteriormente estudiadas. Los resultados fueron los siguientes: Primero, en algunas fórmulas de legibilidad se tendrán que usar técnicas de correlación no lineal. En el análisis de letras, las pruebas F de linealidad mostraron que la mayoría de las regresiones eran curvilíneas. Los resultados en los demás niveles del análisis no fueron concluyentes. Segundo, las fórmulas de legibilidad pueden pronosticar dificultades lo mismo para sujetos en un nivel de habilidad que para sujetos en otros niveles. Tercero, las fórmulas de ligibilidad pueden ser muy útiles para pronosticar las dificultades con palabras individuales, cláusulas independientes y oraciones. Cuarto, la validez de fórmulas de legibilidad que se basan enteramente en las variables linguísticas pueden mejorarse considerablemente. Se encontró una correlación múltiple de .934 entre las variables linguísticas y las dificultades en los pasajes. Quinto, la mejora más notable en el pronóstico de legibilidad resultará del desarrollo de variables linguísticas más sofisticadas. Varias de las nuevas variables exhibieron correlaciones que pasaban de .7 con dificultad de comprensión. Dos de las correlaciones llegaron a .8.

One of the great challenges to scientists of this generation is to learn how to predict and control the difficulty of language. It is almost trite to say that further improvement of public and private life depends upon the ability to transmit ever increasing amounts of knowledge to an increasingly large proportion of the population. But, unfortunately, many adults and children fail to understand what they read, not because the concepts are too difficult or because they lack basic reading skills, but simply because of the complexity of the language in which those concepts are presented. The money wasted annually on materials of this sort mounts into millions of dollars. More appalling is the waste in human terms as, for example, when a child fails to learn and drops out of school, when citizens are unable to inform themselves on important matters of government, and when a worker is unemployed because he cannot keep up with advancing technology. But the cost of lives blighted by poverty, ignorance, crime, frustration, and loss of self respect is not easily calculated. Much of this waste must be directly attributed to the present inability to predict and control language difficulty.

The ultimate objective of this and other similar research is to develop accurate formulas for predicting and controlling the readability of language. But before this objective can be attained, certain basic problems must be solved. This study takes up some of these basic problems. The first is to determine whether various features of writing style are linearly related to the comprehension difficulty of language. The second is to find out if the strengths of these relationships change as a function of the reading abilities of subjects. The third is to find out if useful readability prediction can be made on language units as small as a word or an independent clause. The fourth is to gain some idea of how much accuracy can be attained in passage readability formulas when modern testing and linguistic analysis techniques are used to construct these formulas. The fifth is to try out several new types of linguistic variables (measures of features of writing style) to see if they are useful in predicting language difficulty.

Background

Quality of present readability formulas

It is problematic whether presently available readability formulas help more than they hinder. Because these formulas are easy and inexpensive to apply, they enjoy widespread use by publishers and educators. Publishers use them for "adjusting" the difficulty of instructional materials, and educators use them to decide if instructional materials are suitable for students of a given level of reading ability. Chall (1958) has made a strong case that the formulas are not sufficiently accurate to warrant either of these uses. Their validity correlations range from .5 to only .7, and experiments have shown that they have little, if any, validity when they are used as style guides for "adjusting" the difficulty of materials. Hence, the publishers' "adjustments" of the materials probably do not have the desired effect on the actual difficulty of the materials. But the practice does mislead educators. Since educators use essentially the same formulas as the publishers, they believe that the materials are suitable for their students when, in fact, they are not.

Advances in research techniques

For many years it seemed impossible to improve the formulas because the available research techniques were not equal to the task. In recent years linguists have been making rapid strides in developing objective ways of dealing with language. Also, the cloze test has solved the problem of reliably measuring language difficulty. These new techniques have made it possible to reopen this issue and to deal effectively with many of the basic problems which were beyond the reach of earlier techniques.

The most serious problem encountered in earlier research was the measurement of comprehension difficulty of passages. The best practice available was to give subjects multiple choice tests over the passages. Lorge (1939) criticized this practice because it was uncertain whether the difficulty of the language in the test questions or the difficulty of the language in the passage itself was being measured. The matter was confounded further by the fact that it is notoriously easy to vary the difficulties of these tests simply by changing the alternatives to the question.

The cloze test procedure first conceived by Taylor (1953) was the crucial factor in revitalizing this research. It made possible the accurate measurement of the comprehension difficulty of passages.

A cloze test is made over a passage by replacing every fifth word with an underlined blank of a standard length. Subjects are told to write in the words they think were deleted and responses are scored correct when they exactly match the words deleted. A large amount of research has gone into the development of the cloze procedure. Since Rankin (1965) has written an excellent analysis of this research, only the most relevant studies will be cited. Fletcher (1955) showed that cloze tests are valid measures of comprehension ability. Bormuth (1962) confirmed Fletcher's reports and indicated that cloze tests are valid and highly reliable measures of the comprehension difficulty of passages. MacGinitie (1961) found that cloze items are statistically independent when surrounded by four words of context. Bormuth (1965) found that scoring synonyms correct does not increase the validities of the scores. In every case these results were obtained by administrating the cloze tests untimed and without the subjects first reading the passages from which the tests were made.

Problems studied

If the present study has a single central focus, it is to demonstrate that, by marshaling modern techniques of psychological measurement, linguistics, statistics, and automatic data processing, it is once again possible to make large and useful advances in the study of readability. A range of problems whose solutions remained beyond the reach of earlier research methods has been brought under attack in the present study. These problems were selected: first, because they are basic to achieving accurate readability prediction; and, second, because they represent directions in which research must proceed if the applications of readability are to be raised to the status of a science. These problems by no means represent the only directions in which research must proceed. For example, efforts must be made to discover the basic dimensions in which prose style varies. But the effective study of such problems must await the solution of the problems included in the present study.

Linearity of regressions

Precise readability predictions depend on more than just a knowledge of the sizes of correlations between linguistic variables and comprehension difficulty. Knowledge of whether or not these correlations are linear is required. If they are curvilinear, the shapes of the curves must be determined. Previous investigators have uniformly behaved as if the correlations were linear by using the Pearson product moment correlation model and by using multiple linear regression equations for readability formulas.

While these investigators may have carefully inspected the scatter plots of their correlations for linearity, such an inspection does not necessarily reveal an underlying curvilinearity. These investigators could not measure the difficulties of individual words and clauses. Consequently, they were forced to plot passage difficulty against linguistic variables obtained by averaging values across an entire passage. These averages tend to have distributions that are approximately normal, as can be deduced from the central limits theorem; and two normally distributed variables usually have a linear relationship, if a relationship at all exists between them. While the use of such averages may not violate the correlation models used, it results in the loss of much information that could have been retained by weighting the variable before averaging.

There is evidence that these underlying relationships are not linear. For example, King-Ellison and Jenkins (1954) found a hyperbolic relationship between the Thorndike frequencies of words and the recognition times of subjects. Therefore, there is reason to believe that the precision of readability prediction can be increased by investigating the shapes of these relationships.

Variable strength as a function of reading ability

One of the most troublesome questions in readability prediction is: Do the linguistic variables that influence difficulty for persons at one level of reading achievement have an equal influence on difficulty for students at other levels of achievement? At least three kinds of speculation are possible on this issue. One theory holds that linguistic variables influence difficulty only for persons who have not yet achieved a high degree of skill in reading. If this were true, readability formulas would be applicable only to persons having low reading achievement. A second theory holds that one set of variables has a strong influence on difficulty for persons at one level of reading ability and other sets for persons at other levels. If this theory is correct, a number of readability formulas must be found and each may be used only for persons within a limited range of reading ability. A third theory states that the linguistic variables influencing difficulty for persons at one level of ability have an equal influence on difficulty for persons at other levels of ability. This theory requires the use of only one formula. Should either of the first two theories be proven correct, the consequences for both psycholinguistic theory and readability prediction would be far reaching.

The literature on readability contains statements to the effect that special readability formulas must be developed for use with subjects at different levels of reading ability (Smith & Dechant, 1961). Chall (1958) cited two types of evidence supporting this notion. First, Chall pointed out that the Lorge formula seemed to be better suited for predicting difficulty for young children than the Dale-Chall formula, while the Dale-Chall formula is superior for older children. Second, Chall pointed out that Gray (1935) observed that the sizes of the correlations between linguistic and difficulty variables differed depending on the reading ability of the groups of subjects used. While these effects could also have been attributed to uncontrolled variables in the studies involved, the evidence was sufficiently strong to warrant further investigation of this question.

Readability of small language units

There is an acute need for readability formulas which measure the readabilities of individual words and sentences. Such formulas can be applied extensively to materials such as tiles, indexes, captions, and test items. They may also be used to locate difficult spots within larger texts. However, most readability formulas have been designed only for measuring the readabilities of passages. The reason is not difficult to find. Until the advent of the cloze test there was no practical way to measure the comprehension difficulties of individual words and sentences. While one formula (Forbes & Cottle, 1953) purportedly measures the readability of test items, the manner in which it was constructed makes its value dubious. In the Forbes and Cottle study several readability formulas were applied to each item in a set of test items and these readabilities were averaged for each of the items. The average was then used as the dependent variable. The question investigated in the present study was whether it is possible to obtain useful readability formulas for these smaller language units.

Validities of readability formulas

Probably the most important question raised in the current study is whether it is possible to improve present readability formulas. It is true that formulas currently in use were developed using relatively crude techniques and that they have relatively low validities. At the same time, many of the early investigators made expert use of the techniques available to them. There may be an upper limit on the accuracy with which difficulty can be predicted using just the objectively measurable features of language. If such a limit exists, has it already been reached?

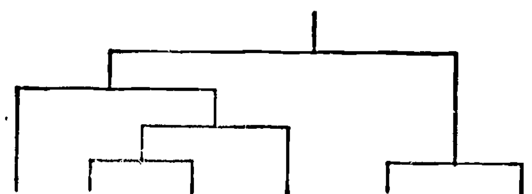
New linguistic variables

Without question the most important advances in readability research should result from the development of new linguistic variables. Linguistic scientists have done much in recent years to improve basic understanding of language and to develop objective ways of describing it. Their techniques have been adapted to the study of psycholinguistic problems and to problems of computer translation of language, resulting in the development of several new and potentially useful variables. Several such variables were investigated. The variables included in the present study by no means exhaust the possible variables that could have been included. The ones included were selected because they are representative of types of variables that could be developed further, if the findings suggest the effort would be worthwhile. These variables and their abbreviations are given below. To conserve space, abbreviations are used in all references to the variables in tables.

Word depth (WOR DEP) The method of deriving word depth was developed by Yngve (1960; 1962). It begins with an analysis and diagraming of the syntactic or immediate constituent structure of a sentence. Then a set of counting rules is applied to the diagram of the structure. Because of the complexity of deriving the variable and because Yngve has defined it, only a general description will be offered here.

Figure 1 shows an example of a sentence which has been analyzed to derive the depths of the words in it. The line diagram shows the syntactic relationships among the words and phrases in this sentence. Note that the horizontal lines at the highest level connect the most distantly related elements while the lines at the lower levels connect the more immediately related constituents.

The counting rules applied to such diagrams are based on several psychological assumptions about the way in which a sentence is produced. First, it is assumed that a sentence is produced starting at the top of this diagram and working downward through the structure until the first word is produced. The remaining grammatical struc-



The dark brown bear sniffed hungrily.
(2) (3) (2) (1) (1) (0)

Total depth, 9; net depth, 3; mean depth, 1.5.

FIG. 1. Illustration of the Word Depth Analysis.

tures and words are then produced from left to right. At each level in this structure there are two elements that must be produced. For example, at the top level both a subject and a predicate must be produced if the sentence is to be grammatically acceptable. Only one of these structures can be expanded and produced at a time and this must be in a "left to right" order. Consequently, if the sentence is to be completed correctly, the speaker must remember that the production of a predicate must follow the production of the subject. Similarly at each lower level in the structure, he must remember to complete any structure which he has begun. The depth of a given word is found by counting the number of structures a speaker has started but not completed at the time that he is producing that word. For example, in Figure 1 the depth of the word *The* is 2 because he must remember that two structures must follow, a noun and a predicate. But the word *dark* has a depth of three because he must remember to produce an adjective, a noun, and a predicate to complete the sentence correctly. Thus, it is assumed in this model that the difficulty of the structure of a sentence derives from the number of grammatical facts that must be stored in the memory as the sentence is produced or interpreted.

The word depth measure can be used in several ways. The depths of the words can be used individually, or they can be totaled across the sentence. In the latter case, total word depth is necessarily related to measures of sentence length, but the necessity of this relationship can be eliminated by subtracting from total word depth the number of words in the sentence. This variable was labeled net word depth (WOR DEP – WOR). The total can also be divided by the number of words in a sentence or a passage to obtain the mean word depth (WOR DEP/WOR).

The word depth variable has been used in only one other readability study (Bormuth, 1964). A correlation of .78 was found between the mean word depths and the comprehension difficulties of a set of passages. When the vocabulary difficulties and mean sentence lengths of the passages were experimentally held constant, the correlation was .77, indicating that the effect of mean word depth is independent of sentence length.

Letter redundancy (LET RED) Carterette and Jones (1963) described a measure of redundancy of passages and prepared a computer program for performing rather lengthy calculations involved in deriving this measure. Again, because of the complexity of deriving this variable only its general description can be given here. A detailed description of the mathematical procedure is given by Carterette and Jones.

The Letter Redundancy variable is a measure of the sequential dependencies between pairs of letters in the words of a passage. Suppose that the letter *m* occurs a number of times in a passage. The frequencies with which this letter is followed by each of the other letters is tabulated and transformed into probabilities, and these probabilities are then used to calculate the redundancy of the passage using the Shannon-Weiner (Shannon, 1943) function. Such a measure of redundancy was calculated for pairs of letters taken four at a time ($LET\ RED_4$) and taken eleven at a time ($LET\ RED_{11}$). The purpose in measuring redundancy across these different numbers of letters was to separate the within word redundancy from the between word redundancy. The former was intended as a type of word difficulty measure, while the latter was intended as a type of syntactic redundancy measure.

Carterette and Jones (1963) calculated LET RED_{11} for samples of text drawn from reading textbooks used at the first, second, third, and fifth grade levels. The measured redundancies of the passages correlated perfectly with the grade levels at which the books were used.

Independent clause frequency (IND CLAU FRE) The *IND CLAU FRE* was derived from data gathered by Strickland (1962) who recorded samples of the language of elementary school children enrolled in grades one through six. Strickland analyzed the grammatical pattern of each independent clause, classified it according to its type, and then tabulated the frequency with which each clause pattern occurred.

The independent clauses in the passages used in the present study were classified using Strickland's scheme. Each independent clause was then assigned a number corresponding to the frequency with which that pattern had been observed in Strickland's language sample. The frequencies for the samples of language taken from children enrolled in grades one, three, and six were combined and used to obtain this measure of frequency.

Figure 2 shows an example of a sentence pattern derived by this analysis.

Sentence

We found and caught the brown bear when we heard him crashing
(1) (2) (4) (M_3)

about in the underbrush.
(M_3 continued)

The pattern of this sentence is 1, 2, 4, M_3.

FIG. 2. Example of Independent Clause Pattern Analysis.

The major constituents of the sentence were classified and a symbol assigned to each constituent. The pattern of an independent clause is the ordered sequence of symbols associated with it. It should be noted that subordinate clauses and constituents of other major portions of clauses are left unanalyzed. Hence, this analysis accounts for only the gross aspects of the patterns of independent clauses. This variable was derived individually for each independent clause, and a mean was obtained for each passage (IND CLAU FRE/IND CLAU).

Ruddell (1963) wrote three pairs of passages which were carefully matched for Dale-Chall readability, subject matter content, and writing style, but which differed in that one member of each pair was written with high frequency clause patterns and the other with low frequency patterns. The data seemed to show that the passages written in low frequency patterns were more difficult to understand. Unfortunately, there is no way to estimate the confidence that can be placed in this result, since no account was taken of item sample error (Bormuth, 1965) in the analysis.

Letter counts (LET) Flesch (1948) found that a count of the mean number of syllables per word had a correlation of .66 with comprehension difficulty. He theorized that this count measures the "abstractness" of words. A less esoteric theorist might speculate that the length of a word is the factor affecting its difficulty and that counting syllables in words affords a crude measure of this factor. If length is the crucial factor, a more discriminating measure could be obtained by counting the number of letters in words. This variable can be used to measure the lengths of words, independent clauses, or sentences.

Parts of speech Boder (1940) found that authors vary considerably with respect to the ratio of adjectives to verbs used in their writing. Carroll (1960), extending Boder's work, found that authors also vary with respect to the frequency with which they use a number of other parts of speech. He factor analyzed a number of these variables and found that part of speech variables seemed to measure a broad range of characteristics of style. However, no effort seems to have been made to determine whether these variables are in any way related to the comprehension difficulty of language.

There was some problem in deciding what classification scheme should be used to classify words in the present study. First, there was the problem of whether to use the structural or the semantic properties of words for classifying them. Because of their superior objectivity, the structural properties were chosen as the principal attributes used in classifying. Then the question of what categories would be used arose. Almost any part of speech category can be broken down into a number of sub-categories; the total number of possible sub-categories is very large. Using a large number of categories results in a more complete understanding of the influence of parts of speech on comprehension, but it also produces categories in which words seldom occur. An attenuation of variance among passages results. Therefore, two different sets of categories were used. One set was similar to the eight categories traditionally used. The other included the four form classes defined by Fries (1952) and a fifth group containing all other structural words. Thus a small number of categories was used in each set and some means of assessing the effects of subclasses was provided. Within each set, the categories were put together in all possible ratios to each other.

Table 1 shows the sub-categories used in classifying the words and the major categories into which they were combined.

Table 1 Sub-categories and combined categories of the part of speech analysis

Sub-categories	*Traditional parts of speech*		*Form classes*
	Name	*Abbreviation*	
Noun	Noun	n	1
Personal Pronoun	Pronoun	pn	1
Interrogative Pronoun	Pronoun	pn	s
Demonstrative Pronoun	Pronoun	pn	s
Indefinite Pronoun	Pronoun	pn	1
Conjunctive Pronoun	Pronoun	pn	s
Verb	Verb	v	2
Auxiliary Verb	Verb	v	5
"Do" Verb	Verb	v	s
Negative	Adverb	adv	s
Simple Adverb	Adverb	adv	4
Number Adverb	Adverb	adv	s
Degree Adverb	Adverb	adv	s
Introducer (there)	Adverb	adv	s
Preposition	Preposition	prep	s
Conjunction	Conjunction	conj	s
Descriptive Adjective	Adjective	adj	3
Ordinal Adjective	Adjective	adj	4
Cardinal Adjective	Adjective	adj	3
Article and Determiner	Adjective	adj	s
Nominal Adjective	Adjective	adj	3
Possessive Adjective	Adjective	adj	3
Interjection	Interjection	intj	s

Variables previously studied

Several variables used in a number of earlier studies were included in the present study. Their presence was essential to provide a frame of reference against which to judge the worth of the new variables, and to insure the adequacy of the answers found to the other problems investigated. The abbreviations and descriptions of these variables are given below.

Word frequency index (WOR FRE) The word frequency measure was obtained by assigning an index number ranging from 0 to 52 to each word in each passage used in the study. These index numbers were taken from the combined counts given by Thorndike and Lorge (1944). Zero was assigned to words not observed in the word count studies, and the numbers 51 and 52 were assigned to the two groups of words that occurred most frequently. This variable was used for individual words and also averaged to find the mean frequencies of words in larger language units. Lorge (1948) found a correlation of .51 between difficulties of passages and the mean frequency of the words in the passages.

Syllable counts (SYL) The number of syllables was counted to measure the lengths of words, independent clauses, and sentences. Flesch (1950) found a correlation of .69 between the mean number of syllables per word in passages and the difficulty of the passages. Gray and Leary (1935) found a correlation of .44 between mean sentence length measured in syllables and the comprehension difficulty of passages. A count of the number of syllables in an independent clause or sentence is necessarily related to the number of words in that unit. This dependence was eliminated by subtracting the number of words from the number of syllables in the independent clause or sentence to obtain the net syllables (SYL – WOR). Mean number of syllables per sentence was also used as a variable (SYL/WOR).

Count of words (WOR) Sentence lengths have been most commonly measured by counting the number of words in them. The correlation between the mean number of words per sentence and the difficulty of passages was found to be .29 by Dale and Tyler (1934), .40 by Gray and Leary (1935) and .52 by Flesch (1948). This variable has been the best single measure of the grammatical complexity of sentences and is incorporated in nearly all of the widely used formulas.

Words on the Dale List of 769 Words (DL 769) The *Dale List of 769 Words* was compiled by Dale (1931) from those words which were common to both the *International Kindergarten Union List* and to Thorndike's 1,000 most frequently used words. This variable was applied to passages in the present study by determining what proportion of the running words in a passage appeared on Dale's list. The same word could be counted several times if it recurred within a passage. The correlation between the comprehension difficulty of passages and this variable was found to be .35 by Dale and Tyler (1934), .35 by Gray and Leary (1935) and .61 by Lorge (1948).

Words on the Dale List of 3,000 Familiar Words (DL 3000) The words on the Dale list were found (Dale & Chall, 1948) to be known to 80 per cent of the children in grade four. The list is used, as is the shorter one, to find the percentage of the words in a passage which appear on the list. Dale and Chall found a correlation of .68 between this variable and the difficulties of passages.

Prepositional phrases (PREP PHR) The PREP PHR variable is usually derived by finding the ratio of prepositions to the total number of words in the passage. The correlation between PREP PHR and comprehension difficulty was found to be .35 by both Dale and Tyler (1934) and Gray and Leary (1935) and .43 by Lorge (1948). It is not entirely clear whether the variable measures sentence length or whether it measures some other attribute of writing style since sentences containing prepositional phrases would seemingly be longer than sentences which do not contain them. In this case, finding the proportions of words appearing in prepositional phrases should result in a higher correlation with difficulty than would the proportion of prepositions. To study the question, the proportion of words in prepositional phrases (WOR PREP PHR) was also used as a variable.

Independent clause (IND CLAU) The sentence, as defined by an initial capital letter and a final punctuation mark, has been the traditional unit for analyzing the grammatical complexity of syntactic units. Coleman (1962) found some evidence that the independent clause is probably a more valid unit of analysis, since a sentence like *The boy went to the orchard and he picked some apples.* may actually be responded to as if it were two separate syntactic units. The conjunction *and* may serve roughly the same function, psychologically, as a period. If this is true, measures of syntactic complexity based on the independent clause should result in higher correlations than variables based on the sentence as it is traditionally defined. In order to study this problem, measures of syntactic complexity were found for both independent clauses and sentences.

From the foregoing descriptions of the problems investigated, it should be evident that the purpose of the present study was not to develop readability formulas for immediate use. To have done so at this time would have been poor strategy. In research a single new development often opens new areas for progress. But new developments simultaneously present new problems. Readability research has been visited not by just one new development but by three. The cloze test has been responsible for two. It made possible the development of formulas for predicting the readability of units as small as a word and it made possible obtaining information essential to the application of more powerful and sophisticated mathematical models to the treatment of readability data. The growing field of psycholinguistics is rapidly assimilating the *ad hoc* theories of readability into a more integrated body of theory, joining the fields of psychology, linguistics, and literary style, and resulting in new approaches to the measurement of the various attributes of prose style. As expected, these developments have presented new problems for study which must be ordered in priority and studied in a systematic fashion. The author has attempted to begin this task.

Procedure

The data for the present study were obtained by making cloze tests from 20 passages selected to represent a variety of prose styles. This provided data on the difficulties of 5,181 words, 405 independent clauses, and 365 sentences as well as on the 20 passages. The details of how these data were collected are given in the sections that follow.

Materials

Passages Twenty passages of between 275 and 300 words each were selected from materials used for instructional purposes in the areas of literature, history, geography, biology, and physical science.

The four passages in each area were selected to provide a roughly even distribution in Dale-Chall readability from about the 4.0 to about the 8.0 grade levels of difficulty. Care was taken to use materials that were not readily available to the subjects, and no two passages were by the same author.

Cloze tests Five cloze test forms were made from each passage with each form made by replacing every fifth word with an underlined blank space 15 spaces in length. No deletions were made from the first and last sentences of each passage. Generally, a word was taken as being defined by the white spaces separating it from other words: *don't, U.S.A.*, 2,*182*, and *re-enter* were deleted as single words. In hyphenated words like *self-made* where both parts were free forms, each part was deleted separately. Commas, apostrophes, and hyphens were deleted along with the rest of the word, except where hyphens separated free forms. The five different test forms, forms A through E, were made by deleting words 1, 6, 11, etc., to make form A, words 2, 7, 12, etc., to make form B, and so on through form E. Every word in a passage, excluding those in the first and last sentences, appeared as a deleted item in exactly one form.

Subjects

The subjects in this study comprised the entire enrollment in grades four through eight of the Wasco Union Elementary School District in Wasco, California. The sample was chosen for its similarity to the general population of school children in the nation with respect to standardized achievement test scores, racial and language backgrounds, and parents' occupational status. The subjects were divided into five form groups, labeled A through E. Their scores on the *California Reading Test* were used to match the means and distributions of reading abilities in the groups. The test had been given by the school personnel several weeks prior to the study. The matching procedure was done a grade level at a time to insure that an approximately equal number of subjects from each grade level was in each form group.

Test administration

The testing extended over a period of eleven school days. On the first day the Stanford Achievement Test: Reading, Form J was administered and the subjects were given a short cloze test to acquaint them with this type of test. For the next ten consecutive school days the subjects then took two cloze tests per day. The test periods were about 50 minutes long and no subject was stopped before he had completed the tests. Each subject took one, and only one, form of the cloze test over each passage and all subjects took a test over each of the twenty passages. A subject was assigned to a form group and all subjects in the same form group took the same cloze test form over each passage. The excellent cooperation of the school authorities and teachers made it possible to administer the tests under almost ideal conditions.

Scoring

Final matching of form groups Subjects who failed to take one or more of the cloze tests were dropped from the study. The form groups were then equated in size by randomly discarding cases from the groups with the larger numbers of subjects. There were 135 subjects in each of the five form groups. The final groups were then checked to see if they were still matched in reading ability. The scores on the Stanford reading test were used as the dependent variable. The F ratio for the between form groups variance shown in Table 2 was not significant. An inspection of the means in the cells also showed that the matching remained extremely close. The over-all mean of the group was 5.5 in grade placement scores, and the range was from about 1.4 to about 12.7 in each form group.

Scoring the cloze tests A subject's response was scored as correct when it exactly matched the word deleted to form the item. Misspellings were scored correct if they were intelligible to the scorer and not ambiguous. A few ambiguities arose in the case of homonyms, each of which could grammatically fit the context of the cloze blank, for example *there* and *their*. The error rate in scoring was held at or below 1 error per 500 items by having a scoring supervisor rescore a 10 per cent sample of every group of 100 tests. In point of fact, the error rate seldom reached as high as 1 error per 1,000 items for a scorer. When the errors exceeded the prescribed level, the entire set was rescored. All scorers were trained to read poor handwriting and phonetic spelling.

Table 2 Analysis of reading achievement scores by cloze test form groups and levels of reading ability

Source	*Degrees of freedom*	*Mean squares*	*F ratio*
Form Groups	4	12.77	.07
Ability Levels	4	248,223.18	1,320.37
Interaction	16	96.89	.52
Within	650	188.02	

Dependent variables

Measures of comprehension difficulty (DIF) The difficulty of each word was found by calculating the proportion of subjects responding correctly when that word appeared as a cloze item. This was called word difficulty. Independent clause difficulty was determined by taking the mean of the word difficulties in each independent clause. Sentence difficulty and passage difficulty were also found by calculating the mean of the word difficulties within those units.

Difficulties at ability levels One of the major problems studied required that difficulties be calculated separately for subjects at different levels of ability. Thus, the subjects in each form group were ranked by the size of their total scores on the *Stanford* reading test and then divided into quintiles. The mean grade placement scores of these groups were 3.2, 4.2, 5.1, 6.3, and 8.7. The F ratio for the between ability levels in Table 2 shows that this resulted in the formation of strata that differed markedly in reading ability. The very small F ratio for the interaction of form groups and ability levels shows that groups at the same ability level were highly similar regardless of the form group in which they appeared. It also may be taken as evidence that the form groups were well matched in the distribution of reading ability. The measures of word, independent clause, sentence, and passage difficulty were then calculated separately for the groups of subjects at each of these levels of ability.

Analyses and results

This part of the report has been subdivided into five sections, one for each of the five major problems investigated. The method of analysis, the results, and a brief discussion of the major results accompanies each section.

Linearity of regressions

Method of analysis The purpose of studying this problem was to determine whether linear correlation models are suitable for treating the raw data in readability studies. The F test of linearity (Guilford, 1956) was the method of analysis used. The Pearson product moment correlation was calculated for the regression of each linguistic variable upon the appropriate measure of comprehension difficulty. An estimate of the amount of variance was yielded that could be accounted for by a fitted straight line. An *eta* correlation, a measure of correlation that is independent of the shape of the regression between variables, was also calculated for the same regression. It yielded an estimate of the amount of variance that could be accounted for by a regression line of any shape. The ratio of the two variances, an F ratio, was used to test the hypothesis that there was no difference between the amounts of variance accounted for by the linear and non-linear correlations. In interpreting these F ratios, it must be understood that Pearson product moment correlations can occasionally exceed the *eta* correlations, thereby causing a negative F ratio.

Word level of analysis Table 3 shows the results of the linearity analysis when applied to the regression on word difficulty of the four variables derived for each individual word.

Table 3 Tests of the linearity of the regressions of linguistic variables on comprehension difficulty at the word level of analysis

Linguistic variable	*r*	η	*F* *
SYL	395	426	12.29
LET	498	520	12.26
WOR FRE	286	334	13.42
WOR DEP	002	058	1.34

* With 14 and 5,166 d.f. and F of 1.64, $p < .05$.

Clearly, all but the word depth variable had a markedly curvilinear relationship. In the case of word depth, the correlation was too slight to yield definitive results. The 14 degrees of freedom in this calculation occur because the distribution on the dependent variable was divided into 15 segments in calculating the *eta*.

An examination of the scatter plots of the correlations showed that they were similar in shape. Words containing few syllables or letters and the high frequency words tended to be easiest. Small increases in word length or decreases in frequency resulted in large increases in difficulty at first. A gradual leveling off occurred as these values continued to change. Examination also showed that a ceiling effect may have accounted for a part of the curvilinearity. A number of the words were of maximum difficulty, indicated by the fact that no subject responded correctly when the word appeared as a cloze item. But the curvilinearity was still clearly apparent even when these words were ignored.

Independent clause level of analysis Table 4 shows the results when the linearity analysis was applied to the variables quantified individually for each independent clause. The curvilinearity was significant only for DL 3000. However, in many other cases the F ratios approached a significant level. Of considerable interest is the fact that syllables per word and letters per word at this level of analysis showed little or no tendency toward curvilinearity.

Table 4 Tests of the linearity of the regressions of linguistic variables on comprehension difficulty at the independent clause level of analysis

Linguistic variable	*r*	*η*	*F**	*Linguistic variable*	*r*	*η*	*F**
SYL	−422	.431	.49	DL 3000/WOR	443	.491	2.89
SYL/WOR	−448	.446	−.13	PREP PHR	−323	.351	1.05
SYL-WOR	−484	.501	1.11	WOR DEP PHR/WOR	−210	.271	1.57
LET	−427	.434	.40	ΣWOR DEP	−329	.334	.20
LET/WOR	−553	.547	.47	ΣWOR DEP-WOR	−288	.294	.17
WOR	−351	.359	.35	ΣWOR DEP/WOR	−202	.257	1.34
ΣWOR FRE/WOR	298	.346	1.76	IND CLAU FRE	049	.115	.54
DL 769/WOR	421	.448	1.47				

* With 14, and 309 d.f. and F of 1.96, $p = <.05$.

This result is partially explained by the well known fact that normally distributed variables tend to exhibit a linear relationship with each other where any relationship exists in the first place. While the distributions underlying the variables were sharply skewed at the word level of analysis, they appeared to be approximately normal when the individual items were averaged to form the variables at the independent clause level of analysis. This normalizing effect follows almost as a matter of necessity from the central limits theorem.

Passage level of analysis Table 5 shows the results at the passage level of analysis. In no case did the F ratio reach the .05 level of significance, but, because of the small number of passages, the test was not very powerful at this level. However, there were several indications that curvilinearity may have existed in much of the data. The most obvious indication is the fact that variables which were first quantified at the sentence and independent clause levels and then averaged to form variables at the passage level, showed a marked decrease in the sizes of the F ratios. Also supporting this possibility were

Table 5 Tests of the linearity of the regression of linguistic variables on comprehension difficulty at the passage level of analysis

Linguistic variable	*r*	*η*	*F**	*Linguistic variable*	*r*	*η*	*F**
	Traditional type variables				*Traditional type variables*		
SYL/WOR	−625	−683	.72	WOR PREP PHR/SEN	−724	−703	−.30
SYL/IND CLAU	−800	−783	−.35	ΣWOR FRE/WOR	537	658	1.27
SYL/SEN	−696	−677	−.25	DL 769	676	743	1.07
LET/WOR	−675	−759	1.43	DL 3000	638	788	2.82
LET/IND CLAU	−807	−794	−.28	WOR/IND CLAU	−769	−732	−.61
LET/SEN	−673	−660	−.16	WOR/SEN	−582	−569	−.11
PREP PHR/WOR	−412	−497	.51	ΣWOR DEP/WOR	−546	−500	−.32
PREP PHR/IND CLAU	−724	−755	.53	ΣWOR DEP/SEN	−502	−490	−.08
PREP PHR/SEN	−753	−741	−.21	IND CLAU FRE/IND CLAU	130	313	.45
WOR PREP PHR/WOR	−456	−474	.10	LET RED_4	350	644	2.49
WOR PREP PHR/IND CLAU	−704	−710	.08	LET RED_{11}	−377	−618	1.94
	Form class ratios				*Form class ratios*		
noun/all other, 1/a	489	623	1.22	adverb/adjective, 4/3	457	665	2.10
verb/noun, 2/1	127	237	.21	structural/adjective, 5/3	369	554	1.23
adjective/noun, 3/1	−606	−708	1.34	noun/adverb, 1/4	−144	−467	1.26
adverb/noun, 4/1	098	463	1.30	verb/adverb, 2/4	−117	−513	1.69
structural/noun, 5/1	−659	−676	.21	adjective/adverb, 3/4	−343	−495	.85
noun/verb, 1/2	−190	−300	.30	adverb/all other, 4/A	228	538	1.67
adjective/verb, 3/2	−573	−559	−.12	structural/adverb, 5/4	−216	−430	.85
verb/all other, 2/A	494	492	−.01	noun/structural, 1/5	674	672	−.03
adverb/verb, 4/2	070	544	2.06	verb/structural, 2/5	666	629	−.39
structural/verb, 5/2	−564	−555	−.07	adjective/structural, 3/5	−390	−548	1.05
noun/adjective, 1/3	548	696	1.79				

Table 5 (Continued)

	Form class ratios				*Form class ratios*		
verb/adjective, 2/3	576	671	1.08	adverb/structural, 4/5	366	588	1.62
adjective/all other, 3/A	−561	−650	.93	structural/all other, 5/A	−692	−658	−.40

	Part of speech ratios				*Part of speech ratios*		
Noun/all other	065	407	.97	Noun/preposition	535	458	−.48
Pronoun/noun	232	342	.36	Pronoun/preposition	469	454	−.09
Verb/noun	200	427	.87	Verb/preposition	513	488	−.16
Adverb/noun	214	357	.47	Adverb/preposition	438	476	.22
Preposition/noun	−458	−415	−.23	Preposition/all other	−421	−476	.32
Conjunction/noun	−338	−473	.71	Conjunction/			
Adjective/noun	−363	−516	.91	preposition	−153	−402	.83
Interjection/noun	613	693	1.02	Adjective/preposition	211	494	1.32
Noun/pronoun	−357	−399	.19	Interjection/preposition	550	601	.46
Pronoun/all other	393	394	.01	Noun/conjunction	564	606	.39
Verb/pronoun	−357	−354	−.01	Pronoun/conjunction	805	852	1.41
Adverb/pronoun	−132	−358	.63	Verb/conjunction	727	743	.26
Preposition/pronoun	−442	−478	.21	Adverb/conjunction	663	680	.21
Conjunction/pronoun	−524	−510	−.10	Preposition/conjunction	215	356	.46
Adjective/pronoun	−469	−461	−.04	Conjunction/all other	−441	−516	.49
Interjection/pronoun	619	704	1.11	Adjective/conjunction	304	526	1.27
Noun/verb	−281	−417	.58	Interjection/conjunction	607	685	.95
Pronoun/verb	305	311	.02	Noun/adjective	365	608	1.88
Verb/all other	421	472	.29	Pronoun/adjective	392	391	−.00
Adverb/verb	115	405	.90	Verb/adjective	432	477	.26
Preposition/verb	−457	−535	.54	Adverb/adjective	471	493	.14
Conjunction/verb	−566	−542	−.19	Preposition/adjective	−112	−529	1.86
Interjection/verb	613	694	1.02	Conjunction/adjective	−251	−401	.58
Noun/adverb	−314	−426	.50	Adjective/all other	−430	−489	.35
Pronoun/adverb	−013	−390	.89	Interjection/adjective	582	646	.68
Verb/adverb	−199	−423	.85	Noun/interjection	533	695	1.93
Adverb/all other	334	440	.51	Pronoun/interjection	541	703	2.00
Preposition/adverb	−392	−450	.31	Verb/interjection	532	694	1.92
Conjunction/adverb	−512	−573	.49	Adverb/interjection	552	714	2.08
Adjective/adverb	−435	−490	.33	Preposition/interjection	442	602	1.31
Interjection/adverb	626	714	1.21	Conjunction/interjection	522	685	1.84
				Adjective/interjection	484	646	1.57
				Interjection/all other	598	671	.84

* With 3 and 15 d.f. and F of 3.29 $p < .05$.

the relatively high F ratios associated with DL 3000, LET RED_4, and LET RED_{11}, none of which were averages of variables quantified at lower levels of analysis.

A third source of evidence was that many of the scatter plots gave a distinct visual impression of curvilinearity. Hence, the only conclusion that can be drawn from the analysis of the data at this level of analysis is that curvilinearity may exist in some of the correlations.

A study using a larger number of passages is required to detect it, if it does in fact exist.

Discussion At the word level of analysis, it was clear that the regressions were curvilinear where a relationship existed in the first place. This was not totally unexpected since the distributions of these variables were known to be markedly skewed. Therefore, it is clear that readability formulas designed to predict readabilities of individual words must take these curvatures into account. This can be done most easily by transforming the numbers obtained from linguistic analyses.

Transformations would probably also increase the accuracy of the readability predictions at the independent clause, sentence, and passage levels of analysis. While the F test of linearity showed that these regressions did not differ significantly from linearity, these results were by no means decisive. That is, being unable to prove that the relationships differed significantly from linearity was not identical with proving that they were linear. Indeed, several of the relationships appeared curvilinear on visual inspection, but these effects predominated at the extremes of the distribution. Because there were fewer items at these extremes, the F ratio was little affected by these tendencies. Consequently, it would be ill advised not to carefully inspect scatter plots when developing a readability formula.

The finding that DL 3000 was curvilinear at the independent clause level of analysis and showed a similar tendency at the passage level has a special interest. This is the most heavily weighted variable in the Dale-Chall formula, which is the most widely used of the current formulas. A similar tendency was noted in the DL 769 variable. When taken together, these results may explain why the formulas have been found (Chall, 1958) to be accurate only over limited ranges of materials. To test this hypothesis the readabilities of the passages were calculated by both formulas and plotted against their comprehension difficulties. The results were pronounced S-shaped curves.

Variable strength as a function of reading ability

Method of analysis Subjects in each form group were stratified into five levels of reading ability; and the difficulties of words, independent clauses, and passages were calculated for each ability level separately. Using each linguistic variable in turn, language units were stratified into thirds and a three by five analysis of variance was performed. A significant interaction effect indicates that the linguistic variable has a closer relationship to difficulty for subjects at one level of ability than for subjects at another.

Word level of analysis Table 6 shows the results of the variable strength analyses when they were applied at the word level of analysis.

Table 6 Interaction effects of levels of reading ability and linguistic variables upon words difficulty

Variable	*F ratio for interaction*	*Variable*	*F ratio for interaction*
SYL	12.01*	WOR FRE	12.84
LET	16.10	WOR DEP	.29

* Where $F_{8,25881} > 2.93$, $p < .05$.

The interactions between ability and linguistic variable levels were significant for all of the linguistic variables except word depth. However, it seems likely that the significant interactions were due to ceiling effects rather than to any property of the independent variables themselves.

A plot of the cell means showed that the interactions were due exclusively to low variation among the means of the most difficult third of the words. These means were quite low, and there was a large number of words to which no subject responded correctly when these words appeared as cloze items. The distributions of word difficulties in these cells were sharply skewed. The main effects in all analyses were significant.

Word depth failed to show a significant interaction, probably due to the fact that its lower correlation with difficulty resulted in the words with maximum difficulty being somewhat more evenly distributed among the cells.

Independent clause level of analysis Table 7 shows the results of the above analysis applied to individual independent clauses. In the independent clause analysis there was no question of ceiling effects distorting the results. The difficulties of the independent clauses ranged from .08 to .68 and appeared to approximate a normal distribution.

Table 7 Interaction effects of levels of reading ability and linguistic variables upon independent clause difficulty

Variable	*F ratio for interaction*	*Variable*	*F ratio for interaction*
SYL	1.74*	DL 3000	1.30
SYL/WOR	.63	PREP PHR/WOR	1.34
SYL-WOR	1.33	WOR PREP PHR/WOR	1.31
LET	.74	WOR DEP	.48
LET/WOR	1.76	WOR DEP-WOR	.45
WOR	.66	WOR DEP/WOR	1.44
WOR FRE/WOR	.74	IND CLAU FRE	5.62
DL 769	1.68		

* Where $F_{8,2011} > 2.93$, $p < .05$.

Only independent clause frequency exhibited a significant interaction with reading ability. The main effects were significant for all variables.

Passage level of difficulty Table 8 shows the results of analysis performed on some of the variables at the passage level of analysis. Because of the labor involved in calculations and because of the regularity in the results, the analysis was applied to only two of the part of speech ratios. None of the interactions at this level of analysis were significant, but the main effects associated with all of the independent variables were significant.

Table 8 Interaction effects of levels of reading ability and linguistic variables upon passage difficulty

Variable	*F ratio for interaction*	*Variable*	*F ratio for interaction*
SYL/WOR	1.40*	WOR FRE/WOR	.43
SYL/IND CLAU	.49	DL 769/WOR	1.41
SYL/SEN	.72	DL 3000/WOR	1.43
LET/WOR	1.37	WOR/IND CLAU	1.37
LET/IND CLAU	1.41	WOR/SEN	1.36
LET/SEN	1.72	WOR DEP/WOR	.54
PREP PHR/WOR	1.20	WOR DEP/SEN	.37
PREP PHR/IND CLAU	.59	IND CLAU FRE/IND CLAU	2.14
PREP PHR/SEN	.49	LET RED_4	1.19
WOR PREP PHR/WOR	1.28	LET RED_{11}	2.20
WOR PREP PHR/IND CLAU	.59	pn/conj	1.26
WOR PREP PHR/SEN	1.44	adj/v	.35

* Where $F_{8,88} > 3.00$, $p < .05$.

Discussion The primary purpose of investigating this problem was to find out if the same readability formulas could be validly used for predicting difficulty for subjects having both high and low reading abilities. Evidently they can. Although the analysis at the word level was indecisive because of ceiling effects, the results at the other levels of analysis showed rather clearly that the correlations between linguistic variables and difficulty do not change as a function of reading ability.

The one exception to this generalization is independent clause frequency. This interaction may have arisen because the variable was based upon the utterances of elementary school children. Such children could be expected to use only a limited range of grammatical patterns, and these would probably be the most frequent patterns in the language. Variations in the frequencies of patterns within the sample of language might be expected to have some correlation with the difficulty of the patterns for young children. But with older children all or most of these patterns would be likely to be over-learned. Hence, this scale in its present form may not extend over a wide enough range to make independent clause frequency of much value for predicting difficulty for any but very young children. In view of its very low correlations with difficulty, it is doubtful if the variable in its present form is of much use even with very young children.

Predicting difficulties of small language units

Purpose In this part of the study one intent was to find out if it is possible to predict the difficulties of individual words, independent clauses, and sentences with enough accuracy to make the development of special formulas of practical interest.

Method of analysis The stepwise multiple regression technique used in this analysis involved adding one variable at a time to the multiple regression equation and computing a new multiple correlation with comprehension difficulty after each new variable was added. Following each step, a partial correlation was calculated between each of the variables not yet added to the equation and comprehension difficulty. The variables already in the equation were held constant in these partial correlations. The new variable added at each step was the one with the highest partial correlation with comprehension difficulty. The procedure was halted when there was no longer a variable having a significant partial correlation. A predicted difficulty was then computed for each language unit in the sample and these were subtracted from the actually observed difficulties to obtain a set of residuals. An inspection of a plot of the residuals reveals curvilinearity in the multiple regression.

It was planned that the analysis would begin with the calculation of a multiple R and the residuals for the data at each level of analysis. If the multiple correlations had been very low and if the plots of the residuals had indicated that curvilinearity was present, appropriate transformations would have been performed on the linguistic variables and the calculations repeated. As indicated below, the multiple correlations were so high that the second step was not necessary.

Word level Table 9 shows the matrix of correlations among the variables derived at the word level of analysis. In interpreting these

Table 9 Intercorrelations among the variables at the word level of analysis

Variable	$\bar{X}$	*s*	2	3	4	5
1. COM DIF	29.66	21.61	−395*	−498	286	002
2. SYL	1.41	.76		807	−487	−082
3. LET	4.47	2.27			−482	−115
4. WOR FRE	46.88	13.17				071
5. WOR DEP	1.65	.76				

* Where d.f. = 5,181 and $r > .026$, $p < .05$.

correlations it should be remembered that none of these variables is linearly related to word difficulty and, in most cases, the linear corre-

Table 10 Stepwise multiple regression analysis of the correlations between linguistic variables and word difficulty

Step	*Variable entered*	*F* to enter*	*R*	R^2	*Increase in* R^2
1	LET	1708.14	.498	.248	.248
2	WOR DEP	21.80	.501	.251	.003
3	WOR FRE	19.70	.504	.254	.003
4	SYL	3.89	.505	.255	.001

* Where F = 3.84, $p < .05$.

lations are probably somewhat lower than the curvilinear correlations among the variables.

The correlations were analyzed using the stepwise multiple regression technique. Table 10 shows the results of that analysis. The multiple correlation with difficulty was .05 when all variables had been entered.

An inspection of the plot of residuals showed that the correlation would have been considerably higher had the linguistic variables been first transformed to make their relationships to word difficulty linear. Of special interest was the behavior of word depth. When letters were held constant, the correlation with word difficulty increased from .002 to .06, suggesting that at the word level of analysis word depth acts as a suppressor variable. Further interpretation is difficult because of lack of linearity in the correlations.

Independent clause level Table 11 shows the correlations among the variables derived at the independent clause level of analysis. While most of these relationships were linear, it is not known how the curvilinearity underlying many of these variables may have affected the sizes of these correlations.

Even so, some interesting relationships appeared. First, the correlations with independent clause difficulty were generally high. All but one, involving independent clause frequency, were significant.

Table 11 Intercorrelations among the variables at the independent clause level of analysis

Variable	$\overline{X}$	s	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1. IND CLAU DIF	32.54	11.98	−42*	−45	−48	−43	−55	−35	30	42	44	−32	−21	−33	−29	−2[illegible]	05
2. SYL	17.99	13.34		35	91	99	28	97	−20	−33	−29	77	30	91	80	27	−07
3. SYL/WOR	1.39	.26			64	29	80	16	−39	−53	−62	18	13	10	05	00	05
4. SYL-WOR	5.30	5.21				88	52	78	−29	−46	−47	68	29	71	60	18	−04
5. LET	56.82	41.88					28	98	−18	−30	−26	76	29	92	82	28	−06
6. LET/WOR	4.42	.70						11	−33	−46	−54	13	09	07	03	−02	09
7. WOR	12.69	8.86							−13	−22	−16	75	28	96	86	30	−08
8. ΣWOR FRE/WOR	47.25	4.47								59	71	−14	−07	−10	−07	−02	−01
9. DL 769/WOR	.71	.17									72	−28	−20	−17	−12	−05	06
10. DL 3000/WOR	.85	.13										−23	−17	−11	−06	01	00
11. PREP PHR	.37	1.51											65	68	57	14	−09
12. WOR PREP PHR	.29	.25												19	11	−04	−04
13. ΣWOR DEP	20.96	18.95													97	50	−08
14. ΣWOR DEP-WOR	8.27	10.88														62	−08
15. ΣWOR DEP/WOR	1.56	.41															−09
16. IND CLAU FRE	162.82	264.77															

* Where d.f. = 405 and $r > .10$, $p < .05$.

The results of the multiple regression analysis of the variables are shown in Table 12. The size of the multiple correlation, .665, exceeded the validity correlations reported for several widely used readability formulas.

Table 12 Stepwise multiple regression analysis of the correlations between linguistic variables and independent clause difficulty

Step	Variable entered	F* to enter	R	R^2	Increase in R^2
1	LET/WOR	177.23	.553	.306	.306
2	WOR	56.25	.625	.391	.085
3	DL 3000/WOR	13.23	.640	.410	.019
4	WOR DEP/WOR	12.83	.655	.429	.019
5	WOR PREP PHR	5.45	.660	.436	.007
6	SYL	4.62	.665	.442	.006

* Where $F = 3.86$, $p < .05$.

In addition, an inspection of the plot of residuals showed that there was a definite tendency toward curvilinearity. This result was not altogether unexpected since DL 3000 had shown a significant degree of curvilinearity in the earlier analyses. On the other hand, the extent of this curvilinearity seemed to be greater than was expected to result from DL 3000 alone. Of major interest also was the large number of variables entering the equation which seemed to suggest that a number of style factors influence comprehension difficulty. Another point of interest was that many of the variables entering the equation were variables being used in this study for the first time. This suggests that future improvements in readability formulas can accrue as much from the development of new variables as from the use of improved analytic and research techniques. Finally, the fact that none of the variables stood out as being of superior importance was important. With slight chance variations in the data, another set of variables could have entered the equation.

Sentence level of analysis Although all of the analyses reported in the present study were performed at both the independent clause and sentence levels of analysis, only the analyses at the independent clause level have been reported because the findings at these two levels were almost identical since over 90 per cent of the sentences were also classed as independent clauses and since almost the identical set of variables was used at both levels. At the sentence level, the multiple R was of interest since it was somewhat higher than at the independent clause level of analysis, .680. Again, the plot of the residuals showed some evidence of curvilinearity.

Discussion The purpose of the analyses was to find out if readability formulas could be designed to make *useful* predictions of difficulty for individual words, independent clauses, and sentences. In the past readability formulas have received widespread use and have had validities ranging from .5 to .7. If these validities may be taken as a criterion of usefulness, then the formulas represented by the multiple correlations in this analysis must be said to be useful. However, it was obvious that the correlations can easily be increased by several methods. First, the curvilinearity of the predictions can be eliminated by performing the appropriate transformations on the variables; and, second, it was evident that the best possible linguistic variables for use at these levels are far from being discovered. Only four variables were explored at the word level and only 15 at the independent clause and sentence levels. The parts of speech, aside from the preposition, have yet to be tried, and it is obvious that other types of variables should be tried.

Validities of readability formulas

Analysis The analysis portion of the study was designed to determine if the use of new linguistic variables and modern testing techniques could improve the ability to predict passage difficulty. No effort was made to transform the linguistic variables or to do any of the other tasks necessary for making predictions with maximum accuracy. Instead, the variables quantified at the passage level were entered directly into a stepwise multiple regression analysis exactly as in the preceding section.

At this point, there were too many variables to be handled simultaneously by the computer program used, so they were analyzed in two sets. The first set consisted of the 47 variables shown in Table 13.

These were the *traditional type* variables numbered 1 to 22 and the form class ratios numbered 23 to 47 in Table 13. Intercorre-

TABLE 13. INTERCORRELATIONS AMONG PASSAGE DIFFICULTY AND THE FORM CLASS AND TRADITIONAL TYPE VARIABLES (N=20)

Name of Variables	X̄	s	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1 PAS DIF	29.66	9.06	-63	-80	-70	-68	-81	-67	-41	-72	-75	-46	-70	-72	54	68
2 SYL/WOR	1.42	0.16		61	28	93	57	22	65	64	50	74	66	51	-47	-87
3 SYL/IND CLAU	21.71	10.23			80	57	100	76	47	93	90	54	92	89	-48	-65
4 SYL/SEN	25.48	14.06				28	82	100	15	64	88	24	65	87	-27	-41
5 LET/WOR	4.47	0.40					55	24	62	59	48	65	58	47	-39	-79
6 LET/IND CLAU	68.20	31.28						79	44	91	91	50	89	88	-45	-61
7 LET/SEN	80.43	45.13							11	60	85	19	59	84	-23	-36
8 PREP PHR/WOR	0.11	0.03								73	58	95	71	55	-44	-67
9 PREP PHR/IND CLAU	1.67	1.00									91	76	98	90	-52	-70
10 PREP PHR/SEN	1.91	1.11										62	90	98	-43	-63
11 WOR PREP PHR/WOR	0.33	0.11											79	65	-52	-72
12 WOR PREP PHR/IND CLAU	5.18	3.36												91	-55	-71
13 WOR PREP PHR/SEN	5.93	3.75													-45	-63
14 ΣWOR FRE/WOR	46.88	2.35														63
15 DL 769	0.69	0.10														
16 DL 3000	0.84	0.09														
17 WOR/SEN	17.92	10.44														
18 WOR/IND CLAU	15.04	6.38														
19 ΣWOR DEP/WOR	1.65	0.23														
20 ΣWOR DEP/SEN	31.45	24.85														
21 IND CLAU FRE/IND CLAU	185.74	93.33														
22 LET RED4	0.34	0.30														
23 LET RED32	0.46	0.26														
24 Noun/all other, 1/A	0.31	0.04														
25 Verb/noun, 2/1	0.47	0.11														
26 Adjective/noun, 3/1	0.31	0.12														
27 Adverb/noun, 4/1	0.11	0.08														
28 Structural/noun, 5/1	1.35	0.24														
29 Noun/verb, 1/2	2.25	0.68														
30 Verb/all other, 2/A	0.14	0.03														
31 Adjective/verb, 3/2	0.71	0.35														
32 Adverb/verb, 4/2	0.23	0.15														
33 Structural/verb, 5/2	3.00	0.90														
34 Noun/adjective, 1/3	3.82	2.15														
35 Verb/adjective, 2/3	1.80	0.99														
36 Adjective/all other, 3/A	0.09	0.03														
37 Adverb/adjective, 4/3	0.43	0.39														
38 Structural/adjective, 5/3	4.87	1.87														
39 Noun/adverb, 1/4	18.80	18.78														
40 Verb/adverb, 2/4	8.01	7.59														
41 Adjective/adverb, 3/4	5.96	6.46														
42 Adverb/all other, 4/A	0.04	0.02														
43 Structural/adverb, 5/4	25.52	28.35														
44 Noun/structural, 1/5	0.77	0.17														
45 Verb/structural, 2/5	0.36	0.09														
46 Adjective/structural, 3/5	0.23	0.07														
47 Adverb/structural, 4/5	0.09	0.07														
48 Structural/all other, 5/A	0.41	0.04														

18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48
-77	-55	-50	13	35	-38	49	13	-61	10	-66	-19	49	-57	07	-56	55	58	-56	46	37	-14	-12	-34	23	-22	67	67	-39	37	-69
44	-08	-03	-07	-24	13	-08	-55	48	-53	25	58	-72	73	-41	72	-52	-68	60	-59	-48	34	25	57	-55	38	-33	-73	47	-58	51
98	48	57	05	-36	24	-33	-29	50	-09	51	37	-59	61	00	65	-49	-57	48	-38	-39	16	10	34	-18	24	-51	-66	32	-28	59
86	81	94	14	-34	23	-48	02	50	19	58	13	-35	44	15	43	-49	-48	42	-22	-37	00	-04	16	05	08	-55	-46	29	-06	48
41	-07	-01	03	-27	11	-07	-46	38	-47	23	49	-61	59	-37	61	-40	-54	48	-57	-33	30	22	43	-52	31	-33	-65	36	-56	49
99	51	61	06	-37	24	-35	-24	47	-05	52	32	-54	56	04	61	-47	-54	45	-36	-36	13	08	29	-14	20	-52	-63	29	-24	58
84	83	96	15	-34	22	-49	08	47	23	57	07	-29	38	18	37	-46	-43	37	-19	-34	-03	-06	11	09	04	-54	-41	25	-03	45
35	-16	-09	21	-17	-09	14	-70	15	-57	16	63	-74	53	-41	74	-11	-34	26	-51	-03	56	46	57	-58	55	-22	-76	09	-61	58
87	32	39	14	-32	12	-17	-50	39	-24	41	53	-72	64	-09	78	-37	-52	42	-42	-28	34	24	45	-30	39	-42	-76	24	-37	62
90	60	71	22	-35	14	-31	-33	46	-07	51	41	-63	61	00	70	-43	-53	45	-37	-32	25	17	37	-18	31	-50	-70	27	-27	61
41	-07	00	19	-21	02	10	-78	28	-54	20	74	-87	68	-36	86	-21	-46	38	-52	-15	57	45	65	-55	59	-26	-85	21	-59	60
85	33	40	15	-32	17	-16	-56	44	-24	40	60	-79	72	-08	84	-39	-56	47	-42	-32	36	25	50	-30	41	-41	-80	30	-36	61
88	61	71	22	-34	18	-29	-38	50	-07	50	48	-69	67	01	75	-44	-56	49	-36	-35	27	17	41	-17	33	-48	-73	32	-26	60
-43	-05	-10	23	57	-29	39	16	-36	06	-58	-09	42	-33	-01	-45	36	39	-29	07	19	-31	-30	-55	07	-48	48	50	-12	11	-60
-51	-12	-14	-01	23	-12	08	54	-46	40	-28	-55	70	-70	28	-71	51	67	-57	41	49	-31	-22	-58	41	-39	30	68	-43	42	-51
-42	-01	01	01	28	-13	14	51	-43	38	-33	-54	71	-68	25	-73	46	60	-50	34	41	-30	-21	-57	37	-39	34	70	-35	38	-54
77	86	99	14	-31	22	-51	17	43	32	57	-03	-19	28	24	27	-41	-36	31	-11	-29	-08	-10	04	18	-02	-51	-31	21	06	39
	59	69	04	-38	26	-40	-15	47	05	55	23	-47	49	11	54	-46	-50	41	-30	-34	08	04	23	-06	16	-53	-57	26	-17	56
		90	-03	-11	30	-48	18	41	46	49	-01	-17	23	40	24	-04	-34	-30	01	-31	-27	-30	-15	32	-22	-45	-25	23	19	27
			14	-29	19	-49	22	39	41	52	-06	-13	23	32	20	-37	-31	27	-03	-27	-15	-17	-05	27	-09	-46	-24	18	15	30
				21	-20	31	-23	-01	-28	-27	17	-07	08	-28	03	28	16	02	-18	11	38	39	27	-27	28	27	04	03	-18	-16
					12	59	-27	-30	-35	-56	24	06	-07	-33	-10	39	27	-16	-14	18	19	15	-04	-28	05	53	19	-09	-19	-34
						-47	07	45	-01	50	-06	-22	31	-03	25	-40	-36	36	-28	-26	06	06	21	-10	15	-51	-35	26	-20	41
							-57	-60	-41	-90	49	-02	-14	-27	-03	71	48	-38	04	42	28	19	-06	-24	09	88	24	-28	-09	-49
								-03	56	34	-94	82	-62	30	-75	-05	29	-25	37	11	-46	-28	-41	47	-39	-28	60	-19	41	-14
									-12	53	06	-43	76	-12	37	-83	-83	96	-48	-84	03	05	49	-22	15	-55	-46	90	-28	30
										19	-48	38	-38	95	-39	-03	17	-29	80	05	-76	-73	-71	97	-69	-09	38	-22	90	-25
											-26	-21	23	08	31	-62	-47	33	-22	-25	03	07	28	01	20	-97	-52	13	-17	80
												-84	69	-27	83	-03	-35	27	-36	-16	36	16	34	-43	30	19	-64	20	-39	19
													-86	18	-95	42	69	-55	48	41	-35	-18	-52	41	-39	27	91	-41	45	-51
														-25	82	-64	-83	87	-57	-72	23	11	56	-42	28	-29	-75	78	-44	35
															-23	-01	11	-25	81	02	-74	-75	-69	96	-68	01	26	-18	90	-27
																-39	-62	47	-49	-31	40	23	52	-43	44	-36	-93	29	-49	65
																	94	-81	36	90	17	19	-30	08	00	67	49	-74	17	-40
																		-87	50	90	-01	06	-44	26	-15	55	68	-79	32	-43
																			-58	-89	10	08	54	-36	19	-39	-51	96	-38	23
																				32	-61	-58	-65	90	-58	36	60	-47	95	-49
																					18	20	-30	09	05	30	32	-91	09	-04
																						97	83	-76	97	-02	-38	-02	-70	41
																							82	-73	95	-05	-25	-03	-68	37
																								-73	90	-26	-53	40	-68	49
																									-70	10	47	-26	97	-39
																										-17	-44	04	-65	51
																											60	-19	30	-83
																												-29	59	-81
																													-24	-03
																														-55

lations were calculated among the variables yielding a rather massive correlation matrix. Data were then entered into the stepwise multiple regression analysis. The results, which are shown in Table 14, were somewhat surprising since the multiple correlation was markedly higher than any that has yet been reported in any known readability study.

Table 14 Stepwise multiple regression analysis of correlations between passage difficulty and the traditional type and form class ratio linguistic variables

Step	*Variable entered*	*F* to enter*	*R*	R^2	*Increase in* R^2
1	LET/IND CLAU	33.61	.807	.651	.651
2	n/str (1/s)	6.04	.862	.743	.092
3	LET/WOR	5.50	.899	.808	.065
4	ΣWOR DEP/WOR	7.59	.934	.872	.064

* When $F > 4.38$, $p < .05$.

A similar analysis was performed on the ratios derived from the counts of words falling into the traditional part of speech categories. The variables are shown in Table 15.

First, a correlation matrix was calculated and then the correlations were used to calculate the multiple regression analysis shown in Table 16.

Table 16 Stepwise multiple regression analysis of correlations between passage difficulty and parts of speech linguistic variables

Step	*Variable entered*	*F* to enter*	*R*	R^2	*Increase in* R^2
1	pn/conj	33.07	.805	.648	.648
2	prep/n	4.27	.845	.719	.071

* When $F > 4.38$, $p < .05$.

Although only one variable actually entered the calculation, the size of the R agreed fairly well with the one obtained from the analysis of other sets of variables. The ratio of prepositions to nouns was allowed to enter this equation in order to gain some idea of how this set of variables would behave if the study were repeated using a larger number of passages.

Discussion It is clearly evident that the present level of precision in readability prediction .n be markedly increased, possibly even over the levels reached by the predictions shown in Tables 14 and 16. Improvements will probably result from the development of more powerful linguistic variables, evidenced by the fact that the variables entering the multiple correlation equations were, almost without exception, newly developed.

It is apparent that available readability formulas contain too few variables and thereby ignore variables of prose style which can make important independent contributions to the multiple correlations.

New linguistic variables

Word depth Word depth seemed a very useful measure. Although its linear correlation at the word level of analysis was not significant, its *eta* correlation at this level, .058, was significant and accounted for a significant amount of variance in the analysis of variance. Further, as seen in Table 9, it appeared to act as a suppressor variable in the multiple correlation with word difficulty. Because of its skewed distribution, the usefulness of the measure could probably be improved by transforming the word depth numbers.

At the independent clause level of analysis, mean word depth appeared to provide a measure of grammatical complexity not closely related to measures of the length of independent clauses. This is indicated by its relatively low correlation with counts of the numbers of letters, syllables, and words in independent clauses shown in Table 11. As a result, this variable contributed substantially to the multiple correlation with independent clause difficulty (Table 12) even after a measure of length was already in the equation. The two other variables based upon word depth exhibited higher correlations with difficulty, but they were also more highly correlated with measures of length.

At the passage level, essentially the same relationships obtained except that mean word depth showed a closer relationship to measures of sentence length than to measures of independent clause length. Mean word depth's correlation with syllables per independent clause was .48; with syllables per sentence, .81; with letters per independent clause, .51; with letters per sentence, .83; with words per independent clause, .59; and with words per sentence, .86. This suggested that mean word depth and measures of sentence length tended to measure grammatical complexity, while measures of independent

TABLE 15a. INTERCORRELATIONS AMONG PASSAGE DIFFICULTY AND EACH OF THE PART OF SPEECH RATIOS. (N=20)

Variables	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32
1	07	23	20	-21	-46	-34	-36	61	-36	39	-36	-13	-44	-52	-47	62	-28	30	42	12	-46	-57	-48	61	-31	-01	-20	33	-39	-51	-44
2		-73	-82	-66	-19	-65	-61	13	70	-57	49	02	58	30	47	12	73	-54	-42	-32	53	-02	34	13	66	-09	23	-47	47	19	37
3			84	59	-16	58	09	09	-68	95	-75	-31	-67	-41	-68	09	-66	89	59	20	-63	-03	-60	09	-54	40	-12	43	-48	-08	-53
4				57	-12	64	15	02	-73	77	-48	-26	-71	-41	-69	02	-86	55	83	07	-79	-12	-75	02	-57	29	-03	41	-49	-12	-51
5					-39	54	31	30	-52	50	-53	49	-56	-21	-47	31	-48	53	28	84	-55	09	-32	30	-86	-41	-74	95	-80	-56	-75
6						-12	42	-63	12	-28	18	-25	31	05	25	-62	08	-18	-30	-35	49	-03	34	-63	33	25	39	-53	54	25	50
7							22	-29	-27	41	-26	08	-28	26	-24	-30	-36	34	29	21	-34	64	-26	-29	-34	10	-14	32	-31	30	-30
8								-31	-11	-13	02	25	-01	01	21	-30	-17	01	-27	27	05	08	40	-31	-28	-28	-24	16	-10	-19	20
9									-15	24	-21	20	-24	-26	-21	100	-09	19	22	37	-33	-34	-23	100	-27	-18	-29	52	-32	-38	-31
10										-75	76	32	98	84	94	-15	92	-67	-72	-22	86	50	78	-15	67	-27	10	-48	59	47	60
11											-83	-40	-76	-55	-80	24	-67	92	67	16	-70	-18	-71	24	-52	48	-07	42	-50	-13	-59
12												25	76	58	79	-21	50	-93	-35	-40	51	17	48	-21	57	-29	27	-51	53	31	60
13													26	39	37	20	31	-27	-42	76	19	34	42	20	-38	-82	-76	58	-40	-38	-30
14														82	95	-25	89	-67	-74	-27	92	47	81	-24	71	-21	17	-55	68	50	68
15															80	-27	74	-50	-58	-06	68	86	65	-26	46	-25	-02	-27	40	62	42
16																-21	83	-70	-78	-18	84	46	89	-21	59	-33	07	-46	57	41	69
17																	-09	19	21	37	-33	-35	-23	100	-28	-18	-29	52	-32	-38	-31
18																		-48	-85	-07	91	48	83	-09	61	-27	-03	-40	51	37	50
19																			36	36	-49	-10	-46	19	-54	36	-24	47	-51	-21	-56
20																				-17	-87	-43	-94	21	-39	37	18	25	-39	-16	-49
21																					-18	14	07	37	-74	-63	-89	90	-72	-62	-65
22																						42	88	-33	66	-15	12	-54	67	42	65
23																							44	-34	18	-13	-15	-03	13	59	14
24																								-23	42	-38	-13	-30	44	23	60
25																									-27	-18	-29	52	-32	-38	-31
26																										38	76	-87	95	72	86
27																											78	-51	40	56	26
28																												-81	79	66	67
29																													-85	-68	-80
30																														69	92
31																															61
32																															
33																															
34																															
35																															
36																															
37																															
38																															
39																															
40																															
41																															
42																															
43																															
44																															
45																															
46																															
47																															
48																															
49																															
50																															
51																															
52																															
53																															
54																															
55																															
56																															
57																															
58																															
59																															
60																															
61																															
62																															
63																															
64																															

TABLE 15b. NAMES OF THE VARIABLES IN TABLE 15a, THEIR MEANS, AND STANDARD DEVIATIONS.

No.	Names of Variables	$\bar{X}$	s
1	Passage Difficulty	26.66	9.06
2	Noun/all other	.26	.05
3	Pronoun/noun	.31	.18
4	Verb/noun	.75	.24
5	Adverb/noun	.23	.14
6	Preposition/noun	.50	.11
7	Conjunction/noun	.29	.18
8	Adjective/noun	.90	.22
9	Interjection/noun	.01	.01
10	Noun/pronoun	.47	3.48
11	Pronoun/all other	.07	.03
12	Verb/pronoun	2.79	.96
13	Adverb/pronoun	.79	.34
14	Preposition/pronoun	2.30	1.90
15	Conjunction/pronoun	1.14	.94
16	Adjective/pronoun	3.95	2.90
17	Interjection/pronoun	.01	.03
18	Noun/verb	1.50	.65
19	Pronoun/verb	.40	.12
20	Verb/all other	.18	.04
21	Adverb/verb	.30	.15
22	Preposition/verb	.76	.38
23	Conjunction/verb	.40	.22
24	Adjective/verb	.32	.57
25	Interjection/verb	.01	.02
26	Noun/adverb	6.25	3.88
27	Pronoun/adverb	1.60	1.02
28	Verb/adverb	4.20	2.21
29	Adverb/all other	.06	.03
30	Preposition/adverb	3.29	2.39
31	Conjunction/adverb	1.59	1.02
32	Adjective/adverb	5.41	3.47
33	Interjection/adverb	.01	.03
34	Noun/preposition	2.15	.88
35	Pronoun/preposition	.69	.51

TABLE 15b. (Continued)

No.	Names of Variables	$\bar{X}$	s
36	Verb/preposition	1.63	.85
37	Adverb/preposition	.53	.50
38	Preposition/all other	.13	.03
39	Conjunction/preposition	.62	.44
40	Adjective/preposition	1.86	.57
41	Interjection/preposition	.02	.07
42	Noun/conjunction	4.64	2.31
43	Pronoun/conjunction	1.30	.79
44	Verb/conjunction	3.26	1.66
45	Adverb/conjunction	1.00	.83
46	Preposition/conjunction	2.32	1.25
47	Conjunction/all other	.07	.03
48	Adjective/conjunction	4.06	2.07
49	Interjection/conjunction	.03	.11
50	Noun/adjective	1.19	.39
52	Pronoun/adjective	.37	.22
52	Verb/adjective	.83	.34
53	Adverb/adjective	.26	.15
54	Preposition/adjective	.58	.16
55	Conjunction/adjective	.33	.19
56	Adjective/all other	.23	.05
57	Interjection/adjective	.01	.02
58	Noun/interjection	3.81	13.39
59	Pronoun/interjection	1.34	4.64
60	Verb/interjection	2.95	10.38
61	Adverb/interjection	1.30	4.45
62	Preposition/interjection	1.76	7.15
63	Conjunction/interjection	.45	1.61
64	Adjective/interjection	3.34	12.63
65	Interjection/all other	.01	.01

33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65
63	54	47	51	44	-42	-15	21	55	56	81	73	66	22	-44	30	61	37	39	43	47	-11	-25	-43	58	53	54	53	55	44	52	48	60
11	21	-41	-31	-29	44	-56	-34	16	49	-06	15	-03	28	-43	17	13	64	-37	-16	-38	45	-39	03	15	-09	-09	-10	-08	-14	-10	-12	14
09	12	79	57	42	-62	56	25	18	-40	28	-07	09	-44	39	-38	09	-10	85	57	52	-21	53	-51	08	06	07	06	07	05	06	06	09
02	06	63	61	38	-62	60	25	02	-43	15	-03	02	-44	45	-39	02	-17	69	66	49	-23	56	-49	02	03	03	03	03	03	03	03	02
31	34	65	62	80	-75	65	66	28	-22	27	03	46	-39	35	-16	30	-31	33	17	88	-67	30	-22	29	24	25	24	25	19	24	22	30
-61	-90	-65	-77	-74	79	-40	-62	-67	-09	-25	-15	-40	52	-14	28	-63	-48	-37	-43	-68	44	-21	45	-65	-11	-12	-11	-14	03	-09	-03	-64
-30	-08	36	31	27	-48	95	23	-25	-87	-47	-69	-43	-78	96	-78	-29	-23	39	29	34	-23	90	-29	-27	-28	-29	-28	-29	-24	-28	-26	-28
-30	-43	-15	-22	-04	03	12	39	-33	-25	-18	-19	-07	09	08	28	-31	-92	-38	-60	-10	-58	-19	75	-32	-03	-04	-03	-05	04	-03	01	-31
100	87	61	70	75	-52	-06	57	98	55	74	65	83	00	-35	28	100	27	27	26	65	-40	-25	-27	99	53	55	53	57	35	50	43	100
-15	-14	-56	-52	-38	56	-30	-24	-13	09	-45	-25	-28	14	-07	06	-15	07	-62	-55	-51	21	-23	41	-14	-14	-14	-14	-14	-12	-14	-13	-14
24	29	84	65	46	-65	44	22	24	-23	45	09	21	-39	25	-33	24	12	95	71	56	-13	47	-65	24	10	10	10	11	05	09	07	24
-21	-21	-64	-43	-42	51	-31	-20	-19	12	-49	-10	-31	21	-13	14	-21	-03	-71	-41	-54	16	-28	45	-20	-18	-18	-18	-19	-15	-18	-15	-20
20	20	-08	04	41	-18	19	44	18	06	-10	-05	36	-09	08	12	20	-29	-45	-45	40	-52	-12	31	19	16	16	16	17	13	15	14	20
-25	-30	-64	-63	-49	68	-35	-33	-23	06	-47	-27	-34	23	-09	11	-24	-05	-67	-62	-61	26	-27	49	-24	-15	-16	-15	-16	-11	-15	-13	-24
-27	-16	-38	-34	-22	29	21	-09	-23	-39	-68	-62	-47	-30	46	-36	-26	-06	-44	-41	-30	06	25	26	-25	-26	-26	-26	-26	-22	-25	-24	-26
-22	-26	-62	-58	-41	57	-30	-10	-20	04	-49	-28	-29	17	-09	18	-21	-22	-74	-73	-56	01	-32	67	-21	-15	-15	-15	-16	-11	-14	-13	-21
100	86	60	70	75	-52	-07	56	97	56	75	66	84	01	-35	29	100	26	27	26	65	-40	-25	-27	99	55	57	55	60	37	53	45	100
-09	-09	-52	-55	-34	55	-36	-25	-08	15	-33	-24	-18	17	-15	10	-09	14	-57	-60	-45	22	-29	38	-08	-09	-10	-09	-10	-08	-09	-09	-09
19	19	74	47	42	-53	36	23	17	-21	44	02	25	-29	19	-22	18	-01	80	43	52	-18	34	-45	18	14	14	14	14	11	13	12	18
21	30	58	68	34	-57	33	10	23	-10	31	26	12	-29	16	-29	22	27	71	88	44	01	40	-69	23	02	03	02	03	-03	02	-01	22
37	35	38	35	70	-51	35	60	34	01	26	06	59	-19	11	06	37	-28	01	-18	75	-64	00	03	35	28	29	28	30	23	28	25	36
-32	-44	-70	-78	-58	80	-46	-44	-33	08	-39	-27	-31	35	-17	20	-33	-11	-66	-72	-65	34	-34	54	-33	-13	-14	-13	-14	-06	-12	-09	-33
-35	-15	-11	-17	-04	02	59	02	-30	-71	-71	-84	-54	-60	79	-65	-34	-12	-13	-24	-06	-04	60	06	-32	-33	-34	-33	-34	-29	-33	-31	-33
-23	-32	-59	-65	-36	56	-33	-03	-23	04	-38	-29	-18	25	-15	29	-23	-40	-76	-91	-50	-11	-42	81	-23	-10	-11	-10	-11	-06	-10	-08	-23
100	87	-61	70	75	-52	-07	57	98	55	74	65	83	00	-35	28	100	27	27	26	65	-40	-25	-27	99	53	55	53	57	35	51	43	100
-28	-28	-55	-55	-65	72	-42	-55	-24	10	-37	-17	-53	27	-15	03	-27	31	-34	-22	-78	62	-16	20	-26	-26	-26	-25	-26	-22	-25	-23	-26
-19	-19	14	-02	-34	13	-02	-44	-16	-18	04	-08	-38	-04	11	-24	-18	28	59	50	-33	54	33	-39	-17	-17	-17	-17	-17	-14	-16	-15	-18
-29	-31	-29	-28	-59	48	-27	-55	-25	-03	-21	-06	-54	18	-07	-07	-28	28	08	23	-66	65	06	-08	-27	-26	-26	-26	-27	-22	-26	-24	-28
52	54	66	68	89	-77	49	71	50	00	42	20	66	-31	15	-04	52	-17	29	17	94	-69	13	-23	51	30	31	30	33	21	29	25	51
-32	-42	-58	-61	-68	80	-44	-56	-30	-08	-34	-14	-52	37	-17	12	-32	08	-39	-31	-81	56	-21	30	-31	-24	-24	-24	-25	-19	-23	-21	-31
-39	-30	-25	-31	-49	36	14	-42	-33	-51	-61	-62	-74	-33	47	-53	-37	17	05	05	-56	49	47	-06	-35	-37	-37	-36	-38	-31	-36	-34	-37
-31	-40	-60	-59	-63	71	-42	-30	-28	06	-38	-17	-50	34	-19	23	-30	-16	-53	-50	-82	28	-32	06	-29	-24	-25	-24	-25	-19	-24	-22	-30
	85	60	69	74	-52	-08	55	97	57	76	66	85	03	-36	30	100	26	27	26	64	-40	-26	-27	99	58	60	58	62	40	56	48	99
		70	82	82	-71	21	65	91	32	49	38	60	-33	-09	-06	88	45	36	40	71	-44	01	-40	89	19	21	19	24	00	17	09	89
			92	83	-86	53	62	63	-08	50	19	45	-50	22	-28	61	15	80	61	81	-46	36	-57	63	16	18	16	20	03	15	09	62
				88	-91	52	69	74	00	47	28	49	-51	19	-26	71	22	65	65	85	-51	32	-57	73	17	18	16	20	01	15	08	72
					-86	50	82	77	09	48	28	66	-40	14	-09	76	04	40	31	95	-69	16	-33	77	26	28	26	30	11	24	17	76
						-67	-76	-53	18	-32	-08	-42	62	-35	33	-53	-08	-61	-54	-86	64	-42	46	-53	-18	-20	-18	-21	-08	-17	-12	-53
							45	-02	-73	-33	-55	-24	-82	90	-74	-06	-11	42	34	54	-39	84	-36	-04	-21	-21	-21	-21	-21	-21	-21	-05
								59	00	26	13	48	-37	11	05	57	-32	08	-06	67	-95	-03	16	58	14	15	14	17	01	12	07	58
									49	67	58	76	-09	-30	19	98	30	28	28	64	-41	-20	-27	100	35	37	35	40	15	32	24	99
										72	91	69	78	-91	84	55	25	-22	-12	02	05	-81	11	52	50	51	50	52	42	49	46	54
											88	87	42	-62	57	73	16	38	28	49	-18	-44	-26	70	64	65	64	66	53	63	58	72
												80	66	-81	76	64	18	06	16	26	-07	-66	-09	61	60	60	60	62	50	59	55	63
													30	-56	57	82	04	13	08	65	-44	-48	-10	79	69	70	68	71	56	67	62	81
														-83	89	-01	-11	-45	-40	-41	28	-81	39	-05	37	37	37	36	41	38	40	-02
															-87	-34	-10	29	24	20	-09	93	-29	-32	-37	-38	-37	-38	-33	-37	-35	-33
																27	-27	-45	-47	-17	-14	-91	53	23	48	48	48	48	46	48	47	26
																	27	27	27	65	-41	-24	-27	100	51	53	51	56	33	49	41	100
																		38	62	09	55	18	-70	29	100	01	00	02	-06	-01	-04	28
																			85	50	07	56	-80	28	08	08	08	09	02	07	04	28
																				44	25	55	-91	28	04	04	04	05	-02	03	00	27
																					-56	26	-49	65	32	33	31	34	19	30	25	65
																						10	-35	-41	-14	-15	-14	-16	-06	-13	-09	-41
																							-59	-22	-30	-30	-30	-30	-27	-30	-28	-23
																								-27	-11	-12	-11	-12	-06	-10	-08	-27
																									43	45	43	48	24	41	33	100
																										100	100	100	98	100	99	48
																											100	100	97	100	99	50
																												100	98	100	99	48
																													97	100	99	53
																														98	100	29
																															100	46
																																38

clause length tended to measure something else, perhaps just length.

Letter redundancy The results for the letter redundancy variables, LET RED_4 and LET RED_{11}, were difficult to interpret. The first problem was that short passages, consisting of only about 1,500 letters each, were used. Carterette and Jones (1963) showed that passages of about 10,000 letters were required to obtain stable estimates of letter redundancy. The second problem was that the Pearson product moment and *eta* correlations with passage difficulty contradicted each other. Table 5 shows that the Pearson product moment correlations were .35 and −.38 for LET RED_4 and LET RED_{11}, respectively. Neither of these was significant. The corresponding *eta* correlations were .64 and −.62, respectively. Both of the latter were significant.

If the *eta* correlations are to be believed, the redundancy variables were among the best used in the present study. Although the F test of linearity failed to demonstrate any curvilinearity, two other sources of evidence strongly suggest that curvilinearity did exist. First, the distributions of the passage redundancy scores were markedly skewed. The mean of LET RED_4 for the total set of passages was .34 while the standard deviation was .30. The mean and standard deviations of LET RED_{11} were .46 and .26, respectively. In both cases skewing was indicated. Second, a visual inspection of the scatter plots of the correlations gave a distinct impression that the relationships were curvilinear.

Independent clause frequency Independent clause frequency appeared to have little, if any, usefulness as a predictor of comprehension difficulty. At the independent clause level of analysis the product moment correlation was not significant. The *eta*, which was .12, is probably not large enough to be of much use as a predictor of difficulty. At the passage level a similarly low correlation was observed. Although visual inspections of the scatter plots gave the impression that both of these relationships may have been curvilinear, they also confirmed that little correlation was present at the start.

Form class and part of speech ratios The part of speech and form class ratios ranged from excellent to poor when judged in terms of the size of their correlations with passage difficulty. Table 17 shows the correlations between each of the form class ratios and passage difficulty. Of the 25 correlations shown in this table, over half were significant at the .05 level. However, one must be cautious about interpreting the sizes of these correlations. Visual inspections of the scatter plots again gave evidence that many of these regressions may have been curvilinear.

Table 17 Correlations between passage difficulty and each of the part of speech ratios

Denominator of the ratio	*Numerator of the ratio*							
	n	*pn*	*v*	*adv*	*prep*	*conj*	*adj*	*intj*
n	—	23	20	21	−46	−34	−36	−61
pn	−36	—	−36	−13	−44	−52	−46	62
v	−28	31	—	12	−46	−57	−48	61
adv	−31	−01	−20	—	−39	−51	−44	63
prep	54	47	51	44	—	−15	21	55
conj	56	81	73	66	22	—	30	61
adj	37	39	43	47	−11	−25	—	58
intj	53	54	53	55	44	52	48	—
all other	07	39	42	33	−42	−44	−43	60

* When $r > .44$, $p < .05$.

Table 18 shows the correlations between the part of speech ratios and passage difficulty. Over half of these correlations were significant. One variable, the ratio of pronouns to conjunctions exhibited one of the highest correlations yet observed for a single variable in a

Table 18 Correlations between passage difficulty and each of form class ratios

Denominators[a] of the ratio	*Numerator of the ratio*				
	n	*v*	*adj*	*adv*	*str*
n	—	13*	−61	10	−66
v	−19	—	−57	07	−56
adj	55	58	—	46	37
adv	−14	−12	−34	—	−22
str	67	67	−39	37	—
all other	49	49	−56	23	−69

[a] Conventional part of speech names were given as an aid in interpretation of this table, but the correspondence of the form class to parts of speech is approximate. The abbreviations used are read: form class 1-n-noun, form class 2-v-verb, form class 3-adj-adjective, form class 4-adv-adverb, and Structural words-Str.

* Where $r > .43$, $p < .05$.

readability study. Nearly all of the ratios containing interjections in either the numerator or the denominator proved to be very good predictors of difficulty.

The contrasts between the correlations in Tables 17 and 18 should also be noted. Some of the form classes contain nearly the same sets of words as some of the part of speech categories. For example, the noun category contained all of the words in form class 1 but form class 1 also contained a few words of other types. This small difference resulted in rather dramatic differences in the sizes of the correlations that the variables exhibited. This is most clearly seen by comparing the sizes of the correlations involving the ratio of nouns to all words and form class 1 to all words. Similar contrasts can be made between form class 2 and the verb, form class 3 and the adjective, and form class 4 and the adverb. These contrasts suggest that subcategories of words can be found which constitute still more powerful predictors of difficulty.

Independent clause as a grammatical unit It appears that measures of length based upon the independent clause yield higher correlations with passage difficulty than those based upon the sentence. The length of these two types of grammatical unit was measured by counting the letters, syllables, or words within them. Table 19 contrasts the difficulty correlations of these measures.

Table 19 Measures of length based on the independent clause and sentence contrasted with respect to the sizes of their correlations with passage difficulty

Unit measured	*Unit of measure*		
	Letter	*Syllable*	*Word*
Independent Clause	−.81	−.80	−.77
Sentence	−.67	−.70	−.58

Unfortunately, there is presently no way to test the significance of the differences between these correlations, since these correlations are, themselves, probably highly correlated. But while it is impossible to test the significance of these differences, it seems highly improbable that differences of this size could have occurred by chance even when only twenty passages are used.

This result may not be as simple as it at first appears; there is some evidence that sentence length and independent clause length may actually measure different aspects of style. The measures of sentence length tended to have higher correlations with the mean word depth variable than the independent clause length measures. When sentence length was measured in letters it was .81; measured in syllables, .83; and measured in words, .86. The corresponding correlations with independent clause length were .48, .51, and .59, respectively. Consequently, independent clause and sentence length measures may actually measure different aspects of prose style.

Units for measuring length The lengths of words, independent clauses, and sentences were measured using as units of measure letters, syllables, and words. The purpose was to find out which units of measure resulted in variables with the highest correlations with passage difficulty. Table 20 shows the resulting correlations.

Table 20 Letters, syllables, and words as alternative units for measuring length, contrasted with respect to correlations with comprehension difficulty

Unit of measure	*Language unit measured*		
	Word	*Independent clause*	*Sentence*
	Word level of analysis		
Letter	−.50		
Syllable	−.40		
	Independent clause level of analysis		
Letter	−.50	−.43	
Syllable	−.45	−.42	
Word		−.45	
	Sentence level of analysis		
Letter	−.42		−.46
Syllable	−.35		−.46
Word			−.39
	Passage level of analysis		
Letter	−.68	−.81	−.67
Syllable	−.80	−.80	−.70
Word	−.77	−.77	−.58

If one may judge from direct comparisons of the correlations in Table 20, it appears that counts of the numbers of letters provide the best measures of word length. Each of these correlations is from .05 to .10 higher than those obtained when syllables provided the unit of measure, and this occurred regardless of the level of analysis upon which the measures were used. When measuring the lengths of independent clauses, it appears to make little, if any, difference what unit of measure is used. On the other hand, it seems that letters and syllables may provide the best units for measuring sentence lengths.

Words in prepositional phrases Table 21 contrasts variables obtained by counting prepositional phrases with variables obtained by counting *words in* prepositional phrases. The contrasts show that somewhat higher correlations were obtained at the independent clause and sentence levels of analysis by using a count of the number of prepositional phrases. At the passage level it seemed to make little difference which method was used to derive the variable.

Table 21 · Number of words in prepositional phrases and number of prepositional phrases contrasted with respect to correlations with comprehension difficulty

Unit measured	*Counts based on–* Words in prepositional phrases	Prepositional phrases
	Independent clause level of analysis	
Independent Clause	–.21	–.32
	Sentence level of analysis	
Independent Clause	–.19	–.37
	Passage level of analysis	
Per Word	–.46	–.41
Per Independent Clause	–.70	–.72
Per Sentence	–.72	–.75

Net measures The net measure is logically neither a measure of the lengths of words nor a measure of the lengths of sentences, but a composite of both. A similar statement may be made about the relationships among measures of net, mean, and total word depths. Table

22 contrasts these three sets of measures at the independent clause and sentence levels of analysis.

Table 22 Net, total, and mean measures contrasted with respect to correlations with comprehension difficulty

Unit of measure	*Measure*		
	Net	*Total*	*Mean*
	Independent clause level of analysis		
Syllable	−.48	−.42	−.45
Word Depth	−.29	−.33	−.20
	Sentence level of analysis		
Syllable	−.51	−.46	−.51
Word Depth	−.29	−.34	−.15

At the independent clause level net number of syllables seems to yield a slightly higher correlation with difficulty than either of the other two measures. Net depth, however, yielded a somewhat higher correlation than the correlation obtained from total depth. Almost exactly the same relationships were obtained at the sentence level of analysis. It would appear from this analysis that the net measures are fairly useful predictors of difficulty at the independent clause and sentence levels of analysis.

Comparisons with previous studies A number of the best variables devised by earlier investigators were included to provide a frame of reference within which to judge the value of the new variables. In this case the value of a variable is judged by the size of its correlation with passage difficulty. In the absence of such variables it would have been uncertain whether the higher correlations exhibited by the new variables were due to some superiority of the cloze technique used in this study or to the quality of the new variable itself.

The correlations between passage difficulty and each of the previously studied variables are shown in Table 23. Table 23 compares the correlations found by earlier investigators with those found in the present study. These correlations were almost identical in size.

Table 23 Correlations between passage difficulty and linguistic variables used in both present and earlier studies

Variable	*Earlier Study* *Author*	*Earlier Study* *Correlation*	*Present Correlation*
SYL/WOR	Flesch (1950)	.69	.63
SYL/SEN	Gray and Leary (1935)	.47	.70
WOR FRE/WOR	Lorge (1948)	.51	.54
PREP PHR/WOR	Lorge (1948)	.43	.46
DL 769	Lorge (1948)	.61	.68
DL 3000	Dale and Chall (1948)	.68	.64
WOR/SEN	Flesch (1948)	.52	.58

Note: Signs of correlations are ignored in this table.

The most notable exception was in the case of syllables per sentence where the two correlations differed markedly. This difference was probably due to the fact that Gray and Leary used very short comprehension tests which may have had low reliabilities.

Discussion The close agreement in the sizes of the correlations found in this and earlier studies provides one basis for judging the values of the new variables. However, one must exercise caution in doing so, since these variables have been compared only with respect to the size of their correlations with passage difficulty. Of at least equal importance are a variable's correlations with other linguistic variables. But this type of analysis is best carried out through multivariate analytic techniques and is beyond the scope of the present study. Hence, the present study must confine itself only to general statements about the magnitudes of the correlations.

Mean word depth evidently measures a form of sentence complexity which is somewhat independent of the lengths of independent clauses and, to a lesser extent, of the lengths of sentences, evidenced by the fact that it entered the multiple regression equation with passage difficulty even after a measure of independent clause length had been entered into the equation. Further, it had a high correlation with comprehension difficulty. These facts are of considerable theoretical and practical importance. From the theoretical standpoint, it would seem that the length of a grammatical unit can be measured separately from its complexity, and both length and complexity seem to contribute independently to the difficulty of written material.

It is likely that the word depth measure can be refined considerably. For example, the present method of analysis forbids making trinary syntactic cuts. Many linguists would argue strongly for making trinary cuts as well. Also, the psychological assumptions underlying the model are stated from the point of view of the difficulty for the writer or speaker. If stated from the point of view of the reader or listener, the prediction power of this variable might be further increased.

The letter redundancy measures seem to be good predictors of comprehension difficulty. While their correlations with difficulty were not uniformly high, this was judged to be due to the facts that the relationships were curvilinear and that only short passages were used.

It seems likely that several useful modifications can be made which may improve this measure. First, it was noted that redundancy calculated across four letters had a negative sign when correlated with difficulty, while redundancy calculated across eleven letters had a positive sign. This might be interpreted as showing that redundancy calculated for eleven letters was confounded and its correlation thereby reduced. If this were true, and if the effects of the first four letters were subtracted, the difficulty correlations of this variable might increase. It should be pointed out that this mathematical model could just as easily be applied to parts of speech to form a measure of grammatical complexity.

Initially the independent clause was thought of as being a more appropriate unit of analysis than the sentence for measuring the lengths of grammatical units; and, indeed, the correlations with difficulty that involved this unit were among the highest observed. What was not expected was that, when the length measures were applied to independent clauses, they seemed to be measuring something different from what was being measured when they were applied to sentences. Because of the relatively close correlation between the sentence length and the mean word depth measures, it was assumed that sentence length was, to some extent, a measure of grammatical complexity, while similar measures applied to independent clauses measured length only.

Other findings were of special interest. The letter appeared to be one of the most useful measures of the length of a unit. This is not really very surprising, since the letter offers a more discriminating metric than either syllables or words. The net measures resulted in variables having fairly high correlations with difficulty at the independent clause and sentence levels. Counting the words in prepositional phrases may have its greatest value at the independent clause level of analysis, but, even here, the results were not substantially superior to those obtained by simply counting the prepositional phrases.

Summary and conclusions

Improvement in the accuracy of predicting and controlling the readability of printed language is vitally needed. Major research in readability almost ceased after 1948. Dormancy occurred because 1] there were no valid methods for measuring the comprehension difficulty of written language, and 2] there was no organized body of basic research and theory upon which readability research could draw. The development of the cloze test solved the problem of validly measuring difficulty while it simultaneously provided additional power and flexibility in making those measurements. The organization of psycholinguistics as a discipline provided readability research with a body of basic research and theory, and linked readability research more closely to research and theory in psychology, linguistics, and literary style.

While these developments opened vast new areas for advancing knowledge of readability, they also revealed certain basic problems which must be solved before sound readability formulas can be developed. The present study investigated what were considered to be most basic questions. These questions and the related results and conclusions from this investigation are summarized as follows.

Are linguistic variables linearly related to comprehension difficulty? If they are not, non-linear regression techniques must replace the linear equations used in present formulas. An F test of linearity was performed to determine if the regression between each linguistic variable and comprehension difficulty was linear. At the word level of analysis, all regressions departed significantly from linearity where a significant correlation existed to begin with. The shapes of the curves obtained suggested that readability formulas designed to predict difficulties of individual words must utilize quadratic equations.

At the independent clause level of analysis, all but the regression involving words on the Dale List of 3,000 words were linear. However, a note of caution must be injected in interpreting these results. First, several of the F ratios approached a significant level. An inspection of these scatter plots showed that, if curvilinearity existed, its effects were most pronounced at the extreme ends of the distributions. Unfortunately, the F tests of linearity is least powerful in detecting curvilinearity at the extremes because so few cases occur there. Hence the failure to detect curvature may have been due to the limitations of present statistical techniques. Second, several variables were averages of one sort or another. It follows almost automatically from the central limits theorem that averages are normally distributed. Since two normally distributed variables almost invariably exhibit a linear relationship, the results of the F tests were more or less predictable. A good illustration was that syllables per word exhibited a curvilinear relationship at the word level of analysis but a linear relationship at the passage level. Consequently, it must be concluded that this problem must receive further investigation at the independent clause level of analysis to see if curvilinearity actually does exist at the extremes and if transforming the variables before averaging or summing them increases the correlations.

At the passage level of analysis, the results were similar except that no F ratio reached significance at the .05 level. That is, inspections of the scatter plots seemed to reveal a tendency for several of the variables to exhibit curvilinearity at the extremes. This was especially evident in the scatter plots of the Dale List and letter redundancy variables. Since only 20 passages were used, the F test had even less power at this level of analysis. Further investigation of the problem was also needed at the passage level of analysis.

Do linguistic variables more strongly influence difficulty for subjects at one level of reading ability than for subjects at other levels? If they do, different readability formulas must be developed for use with subjects at each level of reading ability. A five by three analysis of variance design was used to determine if linguistic variables influence difficulty more strongly for subjects at one level of reading ability than for subjects at other levels. The subjects were stratified into five levels using their scores on the reading achievement test, and the comprehension difficulty of each word, independent clause, and passage was calculated separately for the subjects at each ability level. The language units were then stratified into three levels using their values on a language variable, and the analysis of variance was calculated. The comprehension difficulties of the language units served as the dependent variable. A significant interaction between linguistic and ability levels in this analysis indicates that a variable is a better predictor of difficulty for subjects at one level of ability than for subjects at other levels.

The results show quite clearly that a linguistic variable can usually be expected to predict difficulty equally well for all subjects regardless of their reading levels. This generalization must, of course, be restricted to subjects whose reading achievement scores fall in the range represented in this study, second through twelfth grade in reading achievement grade placement score terms. But when cell means were plotted, no trend was evident which would suggest that this generalization might not hold equally well for subjects of greater reading ability. Only one exception to this generalization, the independent clause frequency variable, was found. However, this was of little interest because of its low correlation with comprehension difficulty. Because of ceiling effects, this problem could not be studied at the word level of analysis. However, this is not a serious limitation since all the variables occurring at the word level of analysis also occurred at the other levels of analysis where they did receive adequate investigation.

Thus, it was concluded that a single readability formula can be used to predict difficulty for subjects at almost any level of reading ability. This greatly simplifies the task of developing readability formulas. Further, it shows that readability analysis should be applied to materials used with students at higher levels of education than has customarily been the case. The amount that even the most able students learn is accurately predictable from the character of the language in their instructional materials.

Can useful readability predictions be made for language units as small as individual words, independent clauses, and sentences? If they can, formulas for making these predictions would be extremely useful for locating difficult parts of texts, and predicting the comprehension difficulties of test items, book indexes, titles, and headings. The objective was to determine if readability formulas designed to predict the difficulties of language units as small as individual words, independent clauses, and sentences have enough validity to warrant their further development. A multiple correlation was found between the linguistic variables and comprehension difficulty at each level of analysis. Had the multiple correlations been low and had the regressions evidenced curvilinearity, non-linear procedures would have been used. The multiple correlation was .51 at the word level of analysis; .67 at the independent clause level; and .68 at the sentence level. While an inspection of the scatter plots showed curvilinearity present, the multiple correlations were high enough to demonstrate the point without further treatment of the data.

Readability formulas found useful in the past had validities ranging from .5 to .7. By these standards, the formulas represented by the regression equations found in the present study must be judged useful. Further, the formulas for predicting the difficulties of small language units can be greatly improved. The most obvious way to improve them is to deal with the problem of curvilinearity. At the word level this will produce a large and easily obtained increase in validity. At the independent clause and sentence levels, correcting for curvilinearity must necessarily be less direct and a large gain in validity is less certain to result. The most important gains in validity will be obtained by developing more valid measures of the attributes of prose style. While this is the most difficult approach, it is the most important, for it will increase our understanding of the psychology of language processes and make important contributions to school curricula in both language composition and comprehension.

Can the validities of readability formulas based entirely upon linguistic variables be improved? If they cannot, research efforts should be diverted to investigate variables which promise to yield gains. Historically, readability prediction formulas utilized only linguistic variables, variables based on objectively measurable features of language, and more or less ignored features that cannot be described objectively. Can the validity of readability formulas based exclusively on linguistic variables be improved beyond the point already reached? The present study with its improved research techniques and large variety of linguistic variables offered an excellent opportunity to study this question. Two multiple regression equations were calculated using the variables at the passage level of analysis. The data were analyzed in two parts because of the large number of variables involved. One multiple correlation was .93; the other, .81. None of the variables used in earlier studies was sufficiently powerful to enter the equations.

It was concluded that much improvement is possible in the validities of readability formulas. The formulas calculated in the present study accounted for 30 to 60 per cent more of the comprehension difficulty variance than was possible with earlier formulas. The increase was due almost entirely to the greater validity of the new linguistic variables. Since these new variables only exemplify (and probably do so only crudely) the range of new variables it is possible to develop, it is not unreasonable to expect carefully constructed readability formulas in the future to exhibit validity correlations closely approaching 1.00.

While this conclusion may at first seem too strong, upon further analysis it is quite reasonable. Readability formulas have *prediction* validity only. The variables in them cannot be interpreted as *causing* the difficulty of text material. Difficulty is undoubtedly *caused* by the language itself, but these linguistic features constitute only a part of the causal variables. Semantic, organizational, and content variables undoubtedly constitute other causes of difficulty. However, linguistic variables can index these other types of variables. For example, discussions of highly complex subject matters are usually accompanied by many conjunctions, while discussions of interpersonal trivia are usually accompanied by the use of personal pronouns. Hence, the linguistic variables reflect the content variables in ordinary writing. Consequently, it is not unreasonable to expect readability formulas based exclusively on linguistic variables to *predict* nearly all of the difficulty variance among texts. But it would be most unreasonable to assert that the linguistic variables in those same formulas *caused* nearly all of the difficulty variance. A clever writer could quickly demonstrate that such a statement is nonsense.

Another point that merits discussion deals with the very large correlations which will be routinely observed in modern readability research. Psychologists have become accustomed to rechecking their calculations for errors when they obtain correlations in excess of .8. This is largely because their measurements usually contain a great deal of error. In readability research, error is much smaller. The linguistic variables are based upon exact counts and contain little or no error. Difficulty measurements are based upon test means which involve only the errors of the means which are much smaller, usually, than the errors of individual test scores. Consequently, readability correlations of .98 and .99 are not impossible and will probably be looked upon as commonplace in the future. As long as readability correlations are not interpreted as implying causation, these high correlations need not cause intellectual indigestion for anyone.

Does the use of various new types of linguistic variables improve the accuracy of readability prediction? If the accuracy of prediction is improved, the underlying rationale of new linguistic variables should be fully exploited to explore and refine them. A large number of previously uninvestigated linguistic variables were included in the present study. Some were representative of fundamentally new approaches to the measurement of prose style. Others were simply refinements of well known variables. The purpose was to determine if variables of these types warranted further development. The methods used were appropriate for exploratory purposes only; the value of a variable was judged almost solely by the size of its correlation with comprehension difficulty. When comparisons were made between variables, tests of significance were not attempted because the nature of the data made tests of significance impossible.

Because of the large number of variables involved, only the most important results can be discussed here. Perhaps of greatest immediate importance was the fact that the length and complexity of a sentence can be measured separately; and, though length and complexity are correlated, each has an independent correlation with difficulty. The extent to which their mutual correlation is necessary is an important problem for linguistic research. Another important finding was that the part of speech variables exhibited very high correlations with difficulty. For example, the ratio between the numbers of pronouns and conjunctions in a passage exhibited a correlation of .81 with difficulty. The third major result was that many of the well known variables were markedly improved by making relatively minor refinements in the way they are derived.

If any general conclusion is warranted by the results from this part of the study, it is that the major advances in readability prediction will result from the development of more sophisticated variables with which to measure the attributes of prose style. Whatever the best possible linguistic variables may be, it is certain that only a beginning has been made in their discovery.

Implications

Readability research can again make rapid strides toward the achievement of its two main goals, readability prediction and readability control, and this research must take new directions. Without question the most important advances should come through the development of better linguistic variables developed through the study of psycholinguistics, linguistics, and literary style.

The wholesale introduction of new variables can have only limited value unless some method is used to integrate each new linguistic variable into a general body of readability theory and unless methods are used whereby large numbers of linguistic variables can be simultaneously evaluated. Factor analysis techniques seem to offer an excellent method of solving both problems. Theory in the area of literary style is poorly developed, is stated in vague subjective terms and contains few statements of how the variables of prose style relate either to each other or to responses in a reader. Factor analysis of linguistic variables can reveal the basic dimensions in which prose style varies among authors, thus providing the theoretical framework within which the properties of new linguistic variables can be studied. The factors or dimensions of prose style can themselves be studied to determine how each influences responses in a reader.

This analytic strategy will greatly facilitate experimental readability research. The ultimate goal of readability research is to gain the ability to control the difficulty of language. This can only be achieved through the use of experimental methods which make it possible to determine if a linguistic variable is causally related to difficulty. At present, it is absurd to attempt to apply experimental techniques to the study of readability because of the great number of variables that must be controlled. Over 150 linguistic variables have been shown to correlate with difficulty. If these variables could be held constant, only a few days' work would be required to devise another 150 variables that would also have to be controlled. Fortunately, many of these variables are closely related to each other. Through factor analytic techniques redundancies can be removed and the entire set of variables reduced to a few basic style factors. This small number of factors can then be easily managed in experimental designs.

The introduction of cloze tests into readability research solved the most urgent and important psychological measurement problem in readability. There are still no adequate instruments for measuring the interest and aesthetic responses passages elicit in subjects. As a result these important aspects of readability have not been dealt with by careful researchers. One possibility that should be explored is to determine whether the principles of the semantic differential can be adapted to construct instruments for these purposes.

REFERENCES

BODER, D. The adjective-verb quotient: a contribution to the psychology of language. *Psychological Record*, 1940, 3, 309–343.

BORMUTH, J. R. Cloze tests as a measure of comprehension ability and readability. Unpublished doctoral dissertation, Indiana University, 1962.

BORMUTH, J. R. Mean word depth as a predictor of comprehension difficulty. *California Journal of Educational Research*, 1964, 25, 226–231.

BORMUTH, J. R. Comparisons among cloze test scoring methods. In J. A. Figurel (Ed.), *Reading and Inquiry*. Newark, Delaware: International Reading Association, 1965.

CARROLL, J. B. Vectors of prose style in language. In T. A. Sebeok (Ed.), *Style in language*. Boston: John Wiley and Sons, 1960, Pp. 283–292.

CARTERETTE, E. C., & JONES, MARGARET H. Redundancy in children's texts. *Science*, 1963, *140*, 1309–1311.

CHALL, JEANNE S. Readability: An appraisal of research and application. *Ohio State University Educational Research Monograph*, 1958, No. 34.

COLEMAN, E. B. Improving comprehension by shortening sentences. *Journal of Applied Psychology*, 1962, 46, 131–134.

DALE, E. A comparison of two word lists. *Educational Records, Bulletin*, 1931, *10*, 484–489.

DALE, E., & TYLER, R. W. A study of factors influencing the difficulty of reading materials for adults of limited reading ability. *Library Quarterly*, 1934, 4, 384–412.

DALE, E., & CHALL, JEANNE S. A formula for predicting readability. *Educational Research Bulletin*, 1948, 27, 11–20.

FLESCH, R. Measuring the level of abstraction. *Journal of Applied Psychology*, 1950, 34, 384–390.

FLESCH, R. A new readability yardstick. *Journal of Applied Psychology*, 1948, 32, 221–233.

FLETCHER, J. E. A study of the relationships between ability to use context as an aid in reading and other verbal abilities. Unpublished doctoral dissertation, University of Washington, 1955.

FORBES, F. W., & COTTLE, W. C. A new method for determining readability of standardized tests. *Journal of Applied Psychology*, 1953, 37, 185–190.

FRIES, C. C. *The structure of English*. New York: Harcourt Brace, 1952.

GRAY, W. S., & LEARY, BERNICE A. *What makes a book readable*, Chicago: University of Chicago Press, 1935.

GUILFORD, J. P. *Fundamental statistics in psychology and education*. New York: McGraw-Hill, 1956.

KING-ELLISON, P., & JENKINS, J. J. The durational threshold of visual recognition as a function of word frequency. *American Journal of Psychology*, 1954, 67, 700–703.

LORGE, I. Predicting reading difficulty of selections for children. *Elementary English Review*, 1939, *16*, 229–233.

LORGE, I. The Lorge and Flesch readability formulas: a correction. *School & Society*, 1948, 67, 141–142.

MAC GINITIE, W. H. Contextual constraint in English prose paragraphs. *Journal of Psychology*, 1961, *51*, 121–130.

RANKIN, E. F. The cloze procedure — a survey of research. In L. T. Thurstone (Ed.), *Yearbook, National Reading Conference*, Milwaukee: National Reading Conference, 1965, Pp. 50–93.

RUDDELL, R. B. An investigation of the effect of the similarity of oral and written patterns of language structure on reading comprehension. Unpublished doctoral dissertation, Indiana University, 1963.

SHANNON, C. E. The mathematical theory of communication. *Bell System Technical Journal*, 1948, 27, 379–423, 623–656.

SMITH, H. P., & DECHANT, E. V. *Psychology in teaching reading*. Englewood Cliffs, N. J.: Prentice-Hall, 1961.

STRICKLAND, RUTH. The language of elementary school children. *Bulletin of the School of Education.* Indiana University, 1962, 34, No. 4.

TAYLOR, W. L. Cloze procedure: a new tool for measuring readability. *Journalism Quarterly,* 1953, 30, 415–533.

THORNDIKE, E. I., & LORGE, I. *The teacher's word book of 30,000 words,* New York: Teachers College, Columbia University, 1944.

YNGVE, V. H. Computer programs for translation. *Scientific American,* 1962, 206, 68–76.

YNGVE, V. H. A model and hypothesis for language structure. *Proceedings of the American Philosophical Association,* 1960, 404, 444–466.