OE FORM 6000, 2/69

DEPARTMENT OF HEALTH, EDUCATION, AND WELFARE
OFFICE OF EDUCATION

ERIC REPORT RESUME

ERIC ACC. NO.				
ED 029 166				
CH ACC. NO.	P.A.	PUBL. DATE	ISSUE	
AA 000 356	24	Mar69	RIEOCT69	

IS DOCUMENT COPYRIGHTED? YES ☐ NO ☐
ERIC REPRODUCTION RELEASE? YES ☐ NO ☐
LEVEL OF AVAILABILITY I ☐ II ☐ III ☐

AUTHOR
Bormuth, John R.

TITLE
Development of Readability Analysis.

SOURCE CODE	INSTITUTION (SOURCE)
JIM14100	Chicago University, Ill.
SP. AG. CODE	SPONSORING AGENCY
RMQ66004	Office of Education, (DHEW), Bureau of Research

EDRS PRICE	CONTRACT NO.	GRANT NO.
0.75;8.60	OEC-3-7-070052-0326	

REPORT NO.	BUREAU NO.
	BR-7-0052

AVAILABILITY

JOURNAL CITATION

DESCRIPTIVE NOTE

170p.

DESCRIPTORS
*Reading Ability; *Reading Comprehension; *Readability; Instructional Materials; *Comprehension Development

IDENTIFIERS

ABSTRACT
The studies reported here are part of a research program whose purpose is to increase the effectiveness with which students acquire knowledge from written instructional materials. The studies had both a basic and an applied objective. The basic objective was to obtain evidence upon which to base a theory of the processes involved in language comprehension. The correlations between a large number of linguistic features and a measure of the difficulty students exhibited in comprehending the written language samples in which those features occurred were determined. The number of linguistic features which can be conceptualized numbers in the hundreds, most of which must be regarded as potentially representing a stimulus involved in the comprehension processes, because present theory of comprehension is too primitive to permit the authors to identify or to rule out more than a few of those features. The applied objective was to develop regression formulas for estimating if instructional materials are suitable for students of varying levels of language comprehension ability. These readability formulas provide a partial solution to the problem of fitting materials to students. That is, students may be provided with materials suited to their levels of comprehension ability not only by manipulating the materials to make them suitably understandable but also by selecting and using just those materials which are suited to the students' comprehension ability. (JL)

GPO 870-390

Final Report

Project No. 7-0052
Contract No. OEC-3-7-070052-0326

DEVELOPMENT OF READABILITY ANALYSES

John R. Bormuth

The University of Chicago

Chicago, Illinois

March, 1969

U.S. DEPARTMENT OF
HEALTH, EDUCATION, AND WELFARE

Office of Education
Bureau of Research

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ACKNOWLEDGEMENTS

A project of the size and complexity of this one cannot be the accomplishment of a single individual. In this project there was a great deal of interaction among all of the personnel, but, nevertheless several contributions may be identified as largely the accomplishments of single individuals.

Peter Menzel, who will shortly join the faculty of Florida State University as an Assistant Professor of Linguistics, was serving as a graduate research assistant on this project when he worked out the sentence structure alternations and the transformational and anaphoric analyses used in these studies. He is publishing (Menzel, 1969) refined versions of these and other analyses.

Mrs. Barbara Dewey assisted in planning and then supervised the administration and handling of the 94,050 cloze test protocols upon which the response data are based. The extremely low attrition of subjects and the high reliability of the data are ample testimony of her foresight, skill, and hard work.

This project involved handling massive amounts of both linguistic and response data, so much in fact, that it could not possibly be performed without making extensive use of the computer for logistic and clerical purposes. Victor Grambsch deserves the credit for designing and programming nearly all of the data handling systems used. Of particular note was his program which performed the complex task of generating numerous linguistic variables from text on which a minimal amount of manual analysis had been performed. John Viesselman helped on that program to the extent that he wrote a routine for assigning Thorndike frequencies to words. Jeffrey Young also made a major programming contribution by devising a program which edited and proofread keypunched text, formed cloze tests and scoring keys over that text, and then cut mimeograph stencils of the tests on a high speed printer in regular upper and lower case type.

Milagros Aquino and Roberta Brockschmidt have each made substantial contributions by administering the office staff and attending to the organization of the project. Miss Brockschmidt supervised the test scoring and Mrs. Aquino (now Dr. Aquino) helped formulate and then supervised the part of speech and Yngve depth analyses.

Special recognition must be given to David Wiley, Associate Professor of Education at the University of Chicago, who advised on research design and data analyses throughout the planning and execution of the project. However, if the project has any faults in design and analysis, they are solely attributable to my failures, not to his.

J.R.B.

INTRODUCTION

Objectives

The studies reported here are part of a research program whose purpose is to increase the effectiveness with which students acquire knowledge from written instructional materials. This problem can be approached in two ways. First, it can be conceptualized as the problem of analyzing the processes involved in language comprehension. The theory which results is then used to design instruction for teaching comprehension skills. Second, it can be conceptualized as the problem of analyzing the linguistic features which cause the language in instructional materials to vary in comprehension difficulty. The theory which results is then used to select and adjust the materials so that students of differing levels of comprehension ability can acquire knowledge from those materials. The studies reported here provide preliminary information essential for the solution of both problems.

The studies had both a basic and an applied objective. The basic objective was to obtain evidence upon which to base a theory of the processes involved in language comprehension. The correlations between a large number of linguistic features and a measure of the difficulty students exhibited in comprehending the written language samples in which those features occurred were determined. The number of linguistic features which can be conceptualized numbers in the hundreds, most of which must be regarded as potentially representing a stimulus involved in the comprehension processes, because present theory of comprehension is too primitive to permit us to identify or to rule out more than a few of those features. To attempt to identify or rule out this many relevant features by experimental methods alone would involve unthinkable costs. But to identify and rule out large numbers of variables on the basis of correlational evidence is both inexpensive and scientifically justified.

The applied objective was to develop regression formulas for estimating if instructional materials are suitable for students of varying levels of language comprehension ability. These readability formulas provide a partial solution to the problem of fitting materials to students. That is, students may be provided with materials suited to their levels of comprehension ability not only by manipulating the materials to make them suitably understandable but also by selecting and using just those materials which are suited to the students' comprehension ability.

A third objective was to establish a criterion level of difficulty for judging the suitability of prose. A readability formula should provide its user with an estimate of whether a passage is suitably understandable for a student having a given level of comprehension ability. Consequently, before a readability formula may be calculated, it is necessary to identify a criterion level of performance which can be regarded as satisfactory. In the past, criteria of suitable performance have been arbitrarily established for interpreting the scores on the various kinds of comprehension tests used to measure the difficulty of passages. In the present studies the effort was directed first at determining for cloze readability tests criterion scores which are comparable to the arbitrary criterion scores accepted in tradition and, second, at establishing empirically a criterion score which has an explicit rational basis for the decision made in its selection.

Problem

A student's ability to comprehend the language in his instructional materials seems to be one of the most basic factors determining the effectiveness of instruction. Much, if not most, of the knowledge contained in the school's curriculum is transmitted through the medium of written language. If a student is unable to understand that language, he will almost certainly fail to learn much of the content of his instruction and both the student and the instruction will fail to attain their objectives.

Only a preliminary effort has been made to estimate the magnitude of this problem.(Bormuth, 1968b). This effort produced the rather grim conclusions that the majority of students comprehend printed language so poorly that they are able to gain little or no information from their instructional materials and that language comprehension instruction in schools seems to have little influence upon improving students' reading comprehension skills. It remains to be seen how this apparent lapse in the schools' program contributes to school drop-out rates, school and cultural disaffection, chronic unemployment, and the like. Logically, at least, they seem connected. If so, the cost of ineffective instruction in language comprehension may be very high indeed.

Neither the analysis of the language comprehension processes nor the analysis of the linguistic variables influencing comprehension have progressed sufficiently to permit instructional problems to be dealt with in an adequate manner. Only within the last decade have psychologists seriously undertaken studies in general language theory. Their studies antedating this period generally exhibit only a crude understanding of the nature of language and a piecemeal approach. But the research occurring during this period, while it demonstrates a somewhat more sophisticated understanding of language, remains piecemeal and unsystematic. A few examples should make these points clear.

First, consider the proposition that early readability studies (Lorge, 1939; Dale and Chall, 1948; and Flesch, 1948 and 1950) showed little understanding of the nature of language. These investigations studied only a limited range of variables. For example, the major readability formulas calculated up to the present time are based on analyses in which the grammatical component of comprehension was characterized only by counts of the number of words, subordinate clauses, prepositional phrases, and pronouns in sentences. This practice ignored the fact that there are many other grammatical structures which almost certainly have to be involved in the comprehension processes. It also ignored the fact that structures such as prepositional phrases probably produce different effects upon the difficulty of sentences depending upon where those phrases are embedded in the sentence structure. Consequently, it is not surprising that available readability formulas have validity coefficients only in the range of .5 to .7 (Chall, 1958),

Next, consider the proposition that the recent research in psycholinguistics has been piecemeal and unsystematic. Aside from a few initial studies (Johnson, 1966; and Ammon, 1968) which demonstrated that language processing responses in humans seem to make use of the phrase structures of sentences, the research has tended to bog down in efforts (Miller, 1962; and Mehler, 1963) to verify the transformational theory of grammar as a model for the sentence comprehension processes. These experiments have tended to focus all attention upon just a narrow range of syntactic structures instead of studying the broad range of structures which exist in language and which are undoubtedly involved in sentence comprehension.

Research Strategy

What seems demanded in order to develop a theory of language comprehension is a strategy broadly enough conceived to examine all of the linguistic variables which are likely to be involved in comprehension but sufficiently economical to be feasible of execution. Because this number of variables is very large, numbering in the hundreds, it is economically difficult to conduct experiments to examine each variable and the interactions among them. The present studies are following a simple strategy which uses correlational and logical procedures to identify and exclude linguistic features which may be involved in the comprehension processes. The object is to reduce the number to a size that can be reasonably dealt with through experimental procedures.

Correlational research designs of the type used here are often condemned with the categorical statement that the existence of a correlation between two variables, say _A_ and _B_, cannot prove that one variable caused the other. This grossly over simplifies the decision theory by which scientific statements are verified. In point of fact, a causal relationship is never _proven_. Rather science proceeds by disproving or ruling out rival theories about what causes a given effect. Specifically, just the following four relationships can hold between any pair of variables:

(1) The variation in _A_ is unrelated to the variation in _B_ so the variables are uncorrelated.

(2) The variation in _A_ causes the variation in _B_.

(3) The variation in _B_ causes the variation in _A_.

(4) The variation in _C_ causes the covariation in _A_ and _B_.

What a theory attempts to explain is which of these relationships holds for a given pair of variables.

The proposition upon which the present studies rest is that the correlational and logical designs used here can achieve almost as much power as experimental studies in identifying which relationship holds for a given pair of variables. Starting with the assumption that the observed variation in two variables is non-random, then the absence of a correlation between the two variables, relationship (1), can be asserted with as much confidence when the design is correlational as when the design is experimental. That is, no causation can be claimed to exist between two variables which do not at least correlate. Next, when two sets of test scores exhibit a correlation, it is often impossible to decide whether A caused B or B caused A, because it cannot be shown that one variable was antecedent in time to the other. But when one variable is clearly the stimulus and the other is the subject's responses, the antecedent variable is identifiable. Finally, correlational designs plus logical analyses can be used to identify or exclude the proposition that C, some alternative variable, caused the covariation observed between A and B. There are usually only a fairly small number of alternative variables which might reasonably be thought to cause the covariation of an A and B. In correlational studies it is quite economical to examine simultaneously a large number of variables, so the likely alternative variables can be examined and many eliminated. One basis for excluding such a variable is that it has a zero correlation with one or both of the other two variables. Another basis for excluding a linguistic variable from further consideration is to show that it is in fact dependent upon some other more general linguistic variable. Both forms of argument were used in these analyses.

ANALYSIS OF LINGUISTIC VARIABLES

Purpose

Objectives: In the studies reported in this section a number of linguistic variables were defined and their correlations with the cloze difficulties of passages were analyzed. The principal object was to furnish a preliminary identification of the linguistic features which serve as stimuli for the reading comprehension processes. Limitations on resources, however, required narrowing the range of linguistic variables in such a manner that only variables defined on the syntax of sentences were intensively studied. Variables defined on vocabulary items and upon intersentence syntax were accorded secondary attention. Also, because these same data were to be used for calculating readability formulas for practical use, some linguistic variables were included for their proven predictive power and ease with which they may be derived, even though they could be ruled out as the variables which could cause prose to vary in difficulty.

Conceptualization of Linguistic Variables: Some manners of conceptualizing linguistic variables yield variables which are more suited to one use than to another. Since this consideration provided the criteria for selecting, defining, and interpreting the variables studied, the principal distinctions will be made explicit. The first distinction is between variables derived pragmatically and variables derived from theoretical considerations. In the earliest readability studies, investigators sometimes examined any variable whatever, regardless of whether that variable could be rationalized even remotely by some reasonable theory. In one study (Gray and Leary, 1935) the investigators even included variables based on the proportions of words in passages which began with various letters. However, no one has since found the correlations observed helpful for explaining language behaviors. Consequently, it seems defensible to study only those variables which may be supported by a reasonable theory and which, therefore, might provide some aid in explaining language comprehension.

A distinction should also be made between manipulable and non-manipulable variables. Variables based upon the semantic referents of words furnish illustrations of non-manipulable variables. Coleman and Miller (1968) devised a variable based on a count of the number of words whose referents are not objectively observable, and Rosenshine (1968) based variables on counts of the number of words referring to the concept of time. Variables of this kind provide little help in constructing a theory of how the difficulty of prose may be manipulated. The referential content of instructional materials must be regarded as unalterable: the author must discuss his topic; to do so he must refer to the relevant concepts; and, therefore, telling him to alter the difficulty of his materials by deleting words referring to a given type of concept is tantamount to telling him that the way to make his topic easier to understand is to write about some other topic. This is not to say that non-manipulable variables should not be studied. Quite the contrary, their study does provide a basis for the theory of the comprehension processes and they present interesting challanges for the theory of language difficulty manipulation. Rather, this is a caveat against interpretations of data which lead to practices as absurd as telling authors that they cannot refer to abstract or time concepts.

Third, a distinction should be made between independent and dependent linguistic variables. It has been customary for investigators to regard variables such as number of letters in a word or number of words in a sentence as the independent variables which cause prose to vary in difficulty. These variables are not directly manipulable. Instead, sentence and word length are, themselves, dependent upon a number of different transformations which can be performed on the language. Hence, it is these transformations which cause variation in language difficulty. Word and sentence length merely constitute variables which happen to be dependent upon the transformations. Again, this is not to argue that dependent linguistic variables should not be studied. Indeed, they often have the valuable property that they summarize the effects of a great many different independent linguistic variables which all have roughly the same effect upon language difficulty. Rather, it is to caution that dependent linguistic variables cannot logically be interpreted as standing in a causal relationship to language difficulty.

Language Sample

A sample consisting of 330 passages, each roughly 110 words in length, was drawn from instructional materials using a two factor stratified random sampling procedure. Nearly all of the passages were verbal text except a few containing some numerals and chemical symbols. In the area of mathematics there were a few passages consisting of story problems, but purely mathematical arguments were excluded from the sample. The stratification procedure used a ten subject matter by five levels of school usage grid. The ten subject matter areas sampled were biology, chemistry, civics, current news, economics, geography, history, literature, mathematics, and physics. The five levels of school usage were grades 1-3, 4-6, 7-9, 10-12, and college, with 7, 8, 8, 7, and 3 passages selected for each of these levels, respectively, within each subject matter category. The proportions at each level represented the author's subjective estimate of the frequency with which the materials were used with students in grades 4 through 12.

The passages were drawn in this manner. First, a list was made of the instructional materials published for the years 1960 to 1966 and each publication was classified into its appropriate subject matter and usage level cell in the sampling matrix. The list was made as exhaustive as possible. Second, when two or more titles written by the same author appeared, all but a randomly chosen one were deleted. Third, when a cell contained fewer titles than was required to permit drawing no more than one passage per title, additional materials were drawn from successively earlier years. Fourth, titles within cells containing more than the required number were randomly discarded to obtain the appropriate number. Fifth, a page and a paragraph number were randomly drawn for each title. Sixth, the materials were opened to that spot and, in general, the passage selected began with that paragraph and ended with the sentence whose termination was closest to 110 words from the beginning. When this procedure obtained passages which depended heavily upon preceding text, which described a graph or figure, or which crossed chapter or major section boundaries, a new passage was drawn from the same materials. The only editing done to the passages was to delete parenthetical and non-restrictive clause references to figures, diagrams, and the like.

Testing Procedure

Cloze Readability Tests: The research establishing the rationale and validity of the cloze readability procedure has been discussed in detail elsewhere (Bormuth, 1967a). That literature, which has now grown rather extensive, strongly supports the statement that tests made, administered, and scored in the manner described here are valid measures of the comprehension difficulty of prose. A cloze readability test is made by deleting every fifth word from a passage and replacing the deleted words with underlined blanks of a uniform length. A hyphenated word is deleted as a unit only if one of its elements represents a bound morpheme as, for example, the co- morpheme in co-chairman.

Numerals are deleted as a unit unless they are spelled out, in which case they are treated like any other word. A student's response to a cloze readability item is scored correct when it exactly matches the word deleted except that misspellings of the correct word are not scored wrong unless they result in the correct spelling of homonyms which would also be grammatically correct in the same context, the verbs bear and bare, for example. Five forms were made of the test over each passage by deleting words 1, 6, 11, etc., to make the first, words 2, 7, 12, etc., to make the second, and so on until all five of the forms possible had been made.

The test booklets were compiled in this fashion: First, the mimeograph stencils containing the tests were placed in five stacks, each stack containing one of the forms of the cloze readability test made from each of the 330 passages. The tests within each stack were then rearranged into a random order, each stack having a different random ordering. The first booklet was then composed of the first 33 tests in the first stack, the next 33 formed the second booklet, and so on until 50 booklets consisting of 33 tests each had been made. Sixty copies of each booklet were printed and a set of printed instructions stapled to the front of each copy. Note that no booklet contained more than one test made from a given passage.

Subjects: Roughly 2,600 students were tested. All were enrolled in school systems in the suburbs of Minneapolis. Local school administrators characterized the population as predominantly white middle class communities. Approximately 500 of the subjects were enrolled in grades 4 through 6, 1000 in grades 7 through 9, and 1000 in grades 10 through 12. These subjects were divided into 50 matched groups using their scores on the California Reading Achievement Test, 1963 edition.

A randomized blocking procedure was used to form the groups. This began by ranking the students at the high school level according to their total grade placement scores on the achievement test. The 50 students ranking highest were then randomly distributed among the 50 groups, one to each group. Then the next highest ranking 50 students were drawn and similarly distributed. This process was repeated until the list of students had been exhausted, and then the entire procedure was repeated for the elementary and junior high school students. However, at the elementary level each student was assigned to a pair of groups and subsequently took two test booklets instead of one. An analysis of the variance, groups by ability blocks, of the achievement test scores resulted in an F of .97 showing that the groups were closely matched in reading achievement.

Test Administration: The reading achievement test was administered approximately two weeks before the cloze testing began. The cloze tests were administered by assigning all the students in the same group to take the tests contained in the same booklet. The elementary grade students took two booklets each but the pair of booklets assigned a student contained no more than one test made from a given passage. The tests were administered under untimed conditions using instructions (Bormuth, 1964) which have come to be accepted as standard for the cloze readability procedure. The testing took place in 40 minute periods spaced daily over a period of roughly two weeks. Testing the elementary grade pupils took about three weeks. Because a few students had moved away or experienced a long term absence during the testing, it was necessary to rematch the groups. This was accomplished by discarding the other 49 students in an ability block in which a student was lost. This reduced the groups to 57 students each but preserved the matching of the groups. The analysis of variance of the achievement scores this time yielded an F of .86.

Reliability of the Data: Because of the massive amounts of data handled, formal quality controls were necessary. To insure correctness in typing the 1650 tests, each passage was keypunched twice and the punches compared using a computer verification program. The test stencils and the test scoring masks were then cut using a computer program written for that purpose. The tests were mimeographed on optical mark scoring forms and the test scorers marked each item right or wrong in the spaces provided on these forms. A 20 per cent sample of protocols for each test form was rescored by a supervisor and the entire set rescored when scoring errors reached 1 in 160 responses. A check of 2 per cent of the protocols after the scoring was completed indicated that the scoring errors ran about 3 per thousand responses. Checks of the accuracy with which the optical mark sensing machine transferred the scoring marks to computer tape revealed no errors.

The reliability of the data was calculated by finding the percentage of students responding correctly to each word when it appeared as a test item, splitting these word difficulties into random halves, averaging the two sets to obtain two estimates of the passage's difficulty, and then correlating across all passages the two sets of means. This correlation was .891 or .942 when corrected using the Spearman-Brown prophecy formula for estimating reliability from comparable split halves.

Passage Difficulty: Passage difficulty was calculated by first calculating the proportion of the students in a group who answered correctly when a word appeared as a cloze item. These word difficulties were then averaged across the passage to obtain the difficulty of the passage. This method was used because it was possible to average the word difficulties across smaller segments of text to obtain the difficulties of individual words, clauses, and sentences and these values were needed in subsequent calculations of the readability formulas.

Linguistic Variables

Vocabulary Variables: The vocabulary variables included were non-manipulable variables selected primarily for their value in interpreting the sentence variables, for their predictive power proven in recent readability studies (Bormuth, 1966; and Coleman and Miller, 1966), or for the ease with which they can be derived. The variable's label, abbreviation, definition, and, if it has not been discussed in one of the recent studies cited, its theoretical support are given below. The abbreviations will hereafter be used to refer to the variables in the tables.

1. Letters per Syllable (LET/SYL)is derived by dividing the number of letters in a word, sentence, or passage by the number of syllables in that same segment of discourse. This variable was defined as a result of an informal observation that, although word length measured in letters provides an excellent prediction of difficulty, some of the most common written words such as though

and <u>bought</u> obtain their length not as a result of actually containing more phonemes but simply because of pecularities of the spelling system. If it is actually phonological length which determines the difficulty of a word rather than graphological length and if word frequency also influences difficulty, then syllable length should correlate positively with passage ease.

2. <u>Letters per Word</u> (LET/W) is derived by dividing the number of letters in a segment of discourse by the number of words that segment contains.

3. <u>Syllables per Word</u> (SYL/W) is obtained by dividing the number of syllables in a segment by the number of words in that segment.

4. <u>Thorndike Frequency</u> (TF/W) is derived by (a) assigning the index numerals 50 and 51 to Thorndike's (1944) <u>A</u> and <u>AA</u> words, respectively, in the <u>G</u> count, (b) assigning each word in a segment its index number from the Thorndike list, reserving zero for words not appearing on the list, (c) subtracting these numbers from 52, (d) summing these index numbers across the segment of text, and (e) dividing by the number of words in the segment. The reason for subtracting the numbers from 52 is merely to obtain smaller and easier sums to work with.

5. <u>Thorndike Frequency of Lexical Words</u> (TFL/WL) is defined by the same procedure except that only counts of lexical words appear in both the numerator and denominator. Lexical words, defined by Fries (1952), consist roughly of nouns, verbs, adjectives and adverbs.

6. Thorndike Frequency of Structural Words (TFS/WS) is defined by the same procedure except that only counts of structural words appear in the numerator and denominator. The structural word categories consist of pronouns, modal and auxiliary verbs, articles and prearticles, prepositions, and so on.

7. Dale Short List Words (DSL/W) is derived by counting the words appearing on the Dale List of 769 Words (Dale, 1931) and dividing by the number of words in that segment.

8. Dale Long List Words (DDL/W) is defined as the number of words appearing on the Dale List of 3000 Easy Words (Dale and Chall, 1948) divided by the number of words in that segment.

Syntactic Structures: The work of the transformational-generative grammarians suggests a number of theories about how the syntax of a sentence might influence its comprehension difficulty. In transformational theory, a sentence, say The horse was ridden by the small boy., is underlain by a deep structure which represents the semantic interpretation of the sentence. In the sentence just given the deep structure might be roughly represented as consisting of the sentences The boy rode the horse., The boy is small. and possibly by a set of operations by which the active sentence The boy rode the horse. is transformed into the passive sentence The horse was ridden by the boy. Hence, sentences as they actually occur i.. speech and writing are regarded as resulting from the transformational operations performed on the deep structures underlying the sentences.

The most obvious theory is one based on the proposition that a structure occurring in a sentence must be traced back to its underlying form before it can be understood. Tracing one structure back to its underlying form may require more operations than tracing a different kind of structure, with the result that the more complex tracing operation increases the likelihood of an error of comprehension. In addition, it might be that some surface structures occur more frequently in the language and that, therefore, some are more familiar and better learned than others. Finally, it is possible for a sentence to contain a complex array of structures, one embedded within the other, thus presenting the reader with a very complex problem in tracing from the surface to the underlying deep structure of the sentence. Variables based upon each of these possibilities were derived.

The analysis upon which these variables were based consisted of identifying the basic structures occurring in English sentences and then counting the number of transformations required to derive the surface structure from the assumed underlying structures. Since a detailed and recently revised version of this analysis is presented elsewhere by Menzel (1969) who developed the analysis for this project, only a structure's label, its abbreviation, the number of transformational operations it involves, and an example will be given for each structure. The numbers in parentheses following the abbbevation indicate the number of transformational operations required to trace the structure back to its underlying form. Since a given structure may take on optional forms this number is given as a range. It should be noted that no strong claim is being made that all of these structures are in fact transformationally derived.

1. Yes-No Questions (YES NO QUES), (1-2); The boy has ridden the horse.-- Has the boy ridden the horse?

2. Wh- Questions (WH QUES), (2-4); The boy has ridden the horse.-- Who has ridden the horse?

3. Tag Questions (TAG QUES), (2-4); The boy has ridden the horse.-- The boy has ridden the horse, hasn't he?

4. Imperative Second Person (IMP SP), (2); You will leave!-- Leave!

5. Imperative Non-Second Person (IMP NSP), (2): We go.-- Let us go.

6. Sentence Negation (SEN NEG), (1-2); Joe saw the man.-- Joe didn't see the man.

7. Constituent Negation (CONST NEG), (1-2); Many people came.-- Not many (or few) people came.

8. Existential There (EXIST THERE), (2); A man is on the corner.-- There is a man on the corner.

9. Cleft It (CLEFT), (3); I saw Joe in the room.-- It was Joe (that) I saw in the room.

10. Anticipatory It (ANT IT), (2-4); To convince him was difficult (for us).-- It was difficult (for us) to convince him.

11. Passive (PASSIV), (3-4); Joe ate the meat.-- The meat was eaten by Joe.

12. Noun Plus Noun Coordinate Structure (CONST CONJ), (3-4); Joe went to the store. + Mary went to the store.→ Joe and Mary went to the store.

13. Sentence Plus Sentence Coordinate Structure (SEN CONJ), (1-5); Joe went home. Mary went home.→ Joe went home and so did Mary.

14. Prenominal Nouns (PRENML N), (3); Joe has a hat. + The hat is for playing baseball.→ Joe has a baseball hat.

15. Prenominal Adjective (PRENML AJ), (2-2); Joe has a hat. + The hat is red.→ Joe has a red hat.

16. Verbal Adjective (VBL AJ), (2-4); Joe saw a dog. + The dog was sleeping.→ Joe saw a sleeping dog.

17. Nominal Adjective (NML AJ), (2); Joe has a light. + The light flashes.→ Joe has a flashing light.

18. Relative Clause (RLTV CLA), (5-6); The man wore the hat. + The man rode a horse.→ The man who rode a horse wore the hat.

19. Prepositional Phrase Adjectival (PREP PH AJVL), (3); The man wore the hat. + The man rode a horse.→ The man with a hat rode the horse.

20. Adverb Derived from an Adjective (DRVD AV), (4-5); Joe entered. + X was slow.→ Joe entered slowly. (Where X stands for the preceeding sentence.)

21. Prepositional Phrase Adverbial (PREP PH AVBL), (3-4); Joe entered. + X was at one o'clock.→ Joe entered at one o'clock.

22. Subordinate Sentence where the Time is the Same (SUB SEN T-S), (2-5); Joe entered. + Bill ate.→ As Joe entered, Bill ate.

23. Subordinate Sentence where the Time is Before (SUB SEN T-B), (2-3); Joe entered. + Bill ate.→ Joe entered before Bill ate.

24. Subordinate Sentence where the Time is After (SUB SEN T-A), (2-3); Joe entered. + Bill ate.→ Joe entered after Bill ate.

25. Subordinate Sentence, Conditional If Clause (SUB SEN IF), (3-7); We hurry. + We will miss the train.→ If we don't hurry, we will miss the train.

26. Tense Shift If Clause (SUB SEN TSH), (3-6); I have fifty dollars. + I will buy a hat.→ If I had fifty dollars, I would buy a hat.

27. Subordinate Sentence, Causal (SUB SEN CAUS), (2-5); The man came. + The man smelled dinner.→ The man came because he smelled dinner.

28. Subordinate Sentence, Purpose (SUB SEN PURP), (2-5); The man came. + The man ate dinner.→ The man came (in order) to eat dinner.

29. Subordinate Sentence, Although (SUB SEN ALTH), (2-4); It rained yesterday. + The children went swimming.→ Although it rained yesterday, the children went swimming.

30. Unequal Comparative (COMP UNEQ), (2-4); Joe has money. + Bill has money.→ Joe has more (or less) money than Bill.

31. Equal Comparative (COMP EQ), (2-4); Joe has money. + Bill has money.→ Joe has as much money as Bill.

32. Superlative Comparative (COMP SUPRL), (3); Joe is a runner. + The runner is → superlative → fast.→ Joe is the fastest runner.

33. Adjective Complement (AJ CMPL), (2-4); He is clever. + He goes.→ He is clever to go.

34. Verb Complement (V CMPL), (2-5); Joe was planning. + Joe started.→ Joe was planning to start.

35. Noun Complement (N CMPL), (2-3); The table is mine. + The table goes.→ The table to go is mine.

36. Factive Nominalization (FACT NOM), (2-3); He came. + X surprised us.→ The fact that he came surprised us.

37. FOR-TO Nominalization (FOR-TO NOM), (3-5); We convinced him. + X was difficult.→ For us to convince him was difficult.

38. Possessive -ING Nominalization (POSS-ING NOM), (3-4); He came. + X surprised us.→ His coming surprised us.

39. Compound Noun 1 (CMPND N 1), (11); The machine is for washing clothes.→ Washing machine, also fire bucket and typing stand.

40. Compound Noun 2 (CMPND N 2), (7); The dumpling is made from potatoes.→ Potato dumpling, also hand loom and steam engine.

41. Compound Noun 3 (CMPND N 3), (6); A person operates the elevator.→ Elevator operator, also truck driver and window cleaner.

42. Compound Noun 4 (CMPND N 4), (6); The man is from China.→ Chinaman, also ranger station and government man.

43. Compound Noun 5 (CMPND N 5), (6); The wound is from a knife.→ Knife wound, also hammer blow and bullet hole.

44. Compound Noun 6 (CMPND N 6), (8); The dichotomy is between form and function.→ Form-function dichotomy.

45. Compound Noun 7 (CMPND N 7), (4-5); The man who dresses like a bat...-- Batman, also White House and dog man.

46. Pronoun (PRONOUN), (3-6); The man-- he.

47. Deleted Nouns (DEL N), (1); The red and green flags are waving. The green flags are torn.-- The green are torn.

48. SOME-ANY in Negatives (S-A NEG), (1); You have some bread.-- You don't have any bread.

49. SOME-ANY in Questions (S-A QUES), (0-1); You have some bread?-- Do you have any (or some) bread?

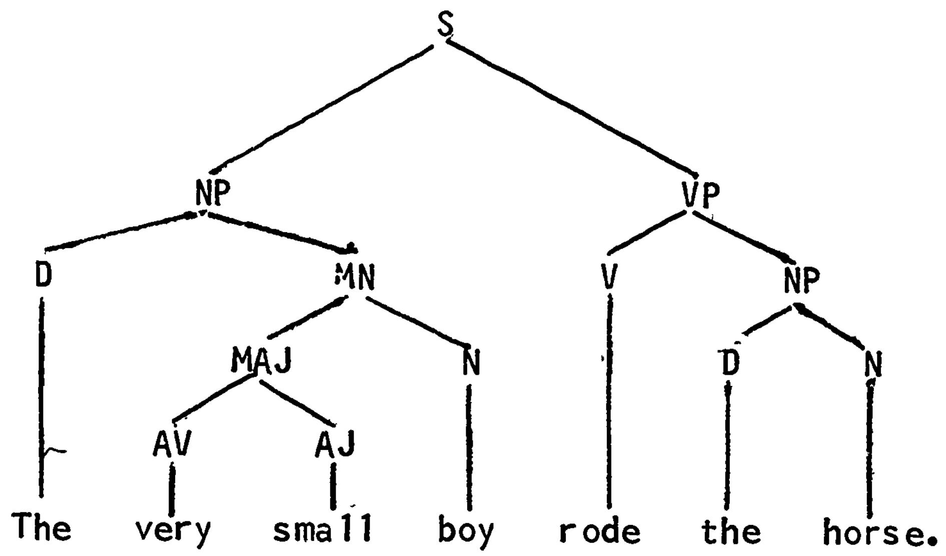
50. Reciprocal (RECIPROCAL), (3); Each man saw the other.-- The men saw each other.

This analysis was used to define the syntactic structure proportion variables. A syntactic structure proportion is derived by counting the number of times a structure occurs in a segment and then dividing by the number of words in that segment. These variables will hereafter be referred to in tables using just the abbreviations given above. If these structures do vary in frequency of use and, therefore, in familiarity or if they differ in the complexity of tracing them back to their underlying interpretations, then the various structures should either increase or decrease the difficulty of the passages in which they occur.

Syntactic Complexity Variables: The complexity of the grammatical structure of a sentence must be studied as a concept distinct from the types or numbers of structures the sentence contains, because, first, measures of sentence complexity correlate with passage difficulty and, second, because complexity can be manipulated somewhat independently of the other two variables. For example, the sentences *The boy who is tall won the race.* and *The tall boy won the race.* are essentially interchangeable semantically and they contain essentially identical underlying forms, but they differ with respect to almost any measure of complexity yet proposed.

The syntactic complexity variables included here are those which are not redundant with variables more easily derived by other means nor which discard much of the structural information about a sentence. To illustrate, Miller (1963) proposed to measure complexity by finding the ratio of the terminal to the total nodes in the phrase structure tree of a sentence. Referring to Sentence (1), a node is a labeled point in

(1)



	The	very	small	boy	rode	the	horse.
Right Depth	2	3	2	1	1	1	0
Left Depth	0	1	2	2	1	2	3

the phrase structure tree drawn over the sentence. Thus, D, NP, and S are nodes. Terminal nodes are just those nodes such as D and AV which immediately dominate words. As it turns out, Miller's ratio can be calculated for all binary branching trees by the function $2(w - 1)/w$, where w is the number of words in the sentence. Since nearly all English sentences are analyzable using binary branching, Miller's variable amounts only to a somewhat circuitous method of counting the number of words in a sentence. Others (Hillel, et al, 1967) have suggested variables which characterize the complexity of a sentence by counting the nodes intervening between the S or top node in the tree and the word having the greatest number of nodes intervening between it and the S node. Sentence (1), for example, has a count of 4. Since procedures of this sort ignore whatever difficulty may arise from the structure of other portions of the sentence, they are regarded, here, as wasteful of information, and therefore, not studied.

The first set of sentence complexity variables, the structure density variables, was obtained by counting the total number of structures (STR) in a segment of prose and then dividing successively by the number of words (W), clauses (CLA), minimal punctuation units (MPU), or sentences (SEN) within the segment to obtain the variables abbreviated, respectively, STR/W, STR/CLA, STR/MPU, and STR/SEN. A minimal punctuation unit is an independent clause plus whatever dependent clauses it might have. These variables measure the density of structures in prose and thereby presumably measure the density of the underlying statements or concepts a reader must interpret. Since it seemed reasonable to argue that the more structures there are in a segment of discourse the greater is the likelihood that errors of comprehension would occur, it also seemed reasonable to expect this variable to correlate with difficulty. The use of the several different denominators arose from the fact that units of discourse such as clauses and sentences are more or less arbitrary divisions making it difficult to select on a priori grounds the theoretically best denominator.

The second set of sentence complexity variables, the transformational complexity variables, was obtained by summing the number of transformational operations associated with each of the structures in a segment of discourse and then dividing by the number of structures, words, clauses, minimal punctuation units, or sentences in the segment. This obtained the variables abbreviated TO/STR, TO/W, TO/CLA, TO/MPU, and TO/SEN. If the comprehension processes do involve tracing structures back to their underlying forms, then each operation involved in those tracing back processes should be accompanied by some likelihood of error. And since these variables measure the density of the operations in a segment of prose, the variables should correlate with the difficulty of the prose.

The third set of variables, called structural complexity, was derived in a somewhat more complicated fashion. A structure occurring in a sentence applies only to a clearly identifiable segment of that sentence. In Sentence (2), for example, the first passive applies to the entire sentence while the other two structures apply only to the last three words. As shown by the numbers under the sentence the last three words

Passive

Relative Clause

Passive

(2) The horse was ridden by the boy who was hurt.

(2)	The	horse	was	ridden	by	the	boy	who	was	hurt.
	1	1	1	1	1	1	1	3	3	3

are affected by three structures and the remaining words by one. Structural complexity is calculated by summing these numbers and dividing by the number of words in the segment. The structural complexity, hereafter abbreviated STR C, of Sentence (2) is 1.6. This variable was included as a different method of quantifying the concept of structure density.

The fourth class of variables are known as Yngve depth. Yngve (1960) developed a model of sentence production which claimed that a person produces sentences by generating the sentence structure tree in a top to bottom direction and from left to right, so that at any given time the speaker has produced only that portion of the left hand side of the tree necessary to produce the word being spoken. That is, as the speaker works down the tree, he produces, so to speak, both branches of a node, but he stores the right hand branch in memory while he proceeds to expand the left hand branch. For example, in Sentence (1) he would have expanded the NP branch of the S node while storing the VP branch, then expanded the D branch while storing in memory the MN branch. Thus, at the time the first word, The, is produced, the nodes MN and VP are being held in memory. Following the production of a word, the last node placed in memory, the <u>MN</u> node in this case, is retrieved and the process resumed. Yngve suggests that the number of nodes in memory at any given point determine the difficulty of producing that portion of the structure. The first row of digits under Sentence (1) gives the depths at each of the words in that sentence. The results of this method of counting will hereafter be called <u>right depth</u>.

Martin (1967) and Bormuth (1966) have each used the same model to predict reader or listener behavior. They based this practice on the reasoning that a word elicits anticipatory responses which the subject stores in memory. Thus, the subject builds, as it were, the tree from left to right and from bottom to top. This produces the same Yngve depth numbers as the model of the speaker. It seems at least as plausible to argue that, for a reader or listener, what should be counted is the number of left branches which the subject must store in order to complete a modification. Specifically, comprehension is here being supposed to take place as a consequence of responses which combine or modify the meanings of the words in a sentence and that those modifications take place in an order which conforms to that determined by the phrase structure of the sentence. Thus, in Sentence (1) some representation of the first word, <u>The</u>, must be stored until the entire structure under the right branch of the <u>NP</u>, the <u>MN</u>, has been processed. Similarly, some representation of the <u>MAJ</u> phrase must be stored until <u>boy</u> has been processed. Thus, what are counted are the numbers of structures remaining incomplete at the time a word is read. For example, at the time boy is read, the MN and NP phrases remain incomplete, giving the sentence a left depth count of two at that point. The bottom row of digits under Sentence (1) show the depths at each of the other words in the sentence. The results from this method of counting will hereafter be referred to as left depth.

Both Martin and Bormuth obtained fairly good predictions of behavior using counts of right depth. However, this could have occurred because there is a necessary correlation between right and left depth counts. The lower limit for either count in binary branching trees is $w - 1$ and the upper limit is $w(w + 1)/2$ where w is the number of words in the sentence. Hence, the effects attributed to right depth may actually have been due to its shared variation with left depth, in which case, it seems necessary to examine both variables simultaneously and in a form which frees them of the necessary correlation due solely to sentence length.

Before the Yngve depth variables can be defined, the nature of the phrase structure analysis upon which the counting operations are based must be considered. Nida (1960) developed a phrase structure analysis which often analyzed phrase structures as discontinuous. For example, in the sentence He called her up., the phrase called up is analyzed as being interrupted by the noun phrase her and is therefore regarded as discontinuous. More modern grammars have avoided showing constituents as discontinuous. While the Nida analysis seems to better represent the underlying deep structure of the sentences, a formal evaluation of the two theories is beyond the scope of the present work. But since Yngve originally based the depth count on the Nida analysis and since basing the counting upon modern phrase structure analyses yields somewhat different results, it seemed necessary to derive variables using both types of phrase structure analysis.

The first set of depth variables defined were obtained by summing the depth number associated with each word and dividing by the number of sentences in the segment. This obtained the variables abbreviated NL/SEN, NR/SEN, ML/SEN, and MR/SEN, where the first variable is interpreted Nida analysis, left depth numbers per sentence and the last is interpreted modern analysis, right depth numbers per sentence. The second set of variables, the <u>net depth</u> variables, were obtained by subtracting the number of words in a sentence from the sum of the depth numbers in that sentence. This obtained the variables abbreviated NET(NL), NET(NR), NET(ML), and NET(MR). <u>Mean depth</u> was obtained by dividing the sum of the depth numbers by the number of words in the segment, thus obtaining the variables abbreviated NL/W, NR/W, ML/W, and MR/W. Finally, <u>relative net depth</u> was obtained in this fashion. A sentence containing $\underline{w}$ words and analyzed by a binary tree can vary in total depth only from $\underline{w} - 1$ to $\underline{w}(\underline{w} + 1)/2$, the maximum range being given by $\underline{w}^2 - \underline{w} + 2$. Hence, where total depth of a particular sentence is $\underline{d}$, the relative net depth of that sentence is $[\underline{d} - (\underline{w} - 1)]/(\underline{w}^2 - \underline{w} + 2)$. These variables were abbreviated RND(NL), RND(NR), RND(ML), and RND(MR).

The <u>syntactic length</u> variables included here are nothing more than a slight elaboration on the conventional practice of measuring the length of sentences. Because there was no way to decide whether to use the clause, minimal punctuation unit, or the sentence as the object being measured or whether to use the letter, syllable, word, clause, or minimal punctuation unit as the unit of measure, all variables possible were derived to obtain the variables abbreviated LET/CLA, LET/MPU, LET/SEN, SYL/CLA, SYL/MPU, SYL/SEN, W/CLA, W/MPU, W/SEN, CLA/MPU, CLA/SEN, and MPU/SEN.

Parts of Speech: An analysis of the parts of speech provides a different method of examining the structures in sentences. For example, in the analysis of syntactic structures, adjectives were examined only when they occurred in a position just before a noun even though the adjective may also occur in other structural positions such as after copulative verbs. This contrast poses the question of whether the influence of adjectives on comprehension results from their adjective-ness regardless of position in a sentence, from the character forced on the structure of the sentence by the use of the adjective, or from some combination of the two. Since the parts of speech are widely discussed in traditional books on grammar, definitions and examples will be provided only for the few special subclasses which are either rarely discussed or were defined specifically for the present investigations.

1. Proper Noun (prp n).

2. Proper Compound Noun (prp cmp n); Proper nouns spelled as one word but made up of two words: Englishman and Rockport.

3. Common Noun (cmn n).

4. Common Compound Noun (cmn cmp n); analogous to proper compound nouns.

5. Numerical Noun (numeric n); all numbers, Arabic or spelled, appearing in noun positions.

6. Gerund (gerund).

7. Infinitive (inf).

8. Personal Pronoun (pers pn).

9. Designative Pronoun (dsgn pn).

10. Compound Pronoun (cmp pn).

11. Adjectival Pronoun (ajvl pn); adjectives with deleted nouns or possessives appearing in noun positions: My favorite is .... or the red is ....

12. Verbal Adjectival Pronoun (vbl ajvl pn); adjectives derived from verbs and appearing in noun positions: The wounded are ....

13. Proper Possessive Pronoun (prp pos pn); proper possessive pronouns appearing in noun positions: Mary's is .... or France's are ....

14. Common Possessive Pronoun (cmn pos pn); analogous to proper possessive pronoun.

15. Personal Possessive Pronoun (pers pos pn).

16. Compound Possessive Pronoun (cmp pos pn).

17. Transitive Active Verb (trn act v).

18. Transitive Passive Verb (trn pas v).

19. Intransitive Verb (intrn v).

20. Linking Verb (link v).

21. Compound Verb (cmp v); analogous to compound nouns.

22. Auxiliary Verb (aux v).

23. Modal Verb (modal v).

24. Pro-Verb (pro v); auxiliary or modal verbs used in place of main verbs: Joe will not ride but Bill will.

25. Verb Contraction (cont v).

26. Infinitive without to (inf w-o to).

27. Article (article).

28. Designative Adjective (dsgn aj); words like this, that, less, all, or enough appearing in adjective positions.

29. Basic Adjective (bas aj); adjectives taking -er and -est suffixes.

30. Derived Adjective (drvd aj); adjectives not taking -er or -est suffixes.

31. Proper Nominal Adjectives(prp aj); proper nouns modifying another noun: Windsor soap, Charles Dickens.

32. Common Nominal Adjective (cmn aj); analogous to proper nominal adjectives.

33. Compound Adjective (cmp aj).

34. Numerical Adjective (numeric aj).

35. Verbal Adjective (vbl aj).

36. Proper Possessive Adjective (prp pos aj).

37. Common Possessive Adjective (cmn pos aj).

38. Prenominal Possessive Adjective (pnml pos aj).

39. Negative Adjective (neg aj); the word no appearing in an adjective position.

40. Common Adverb (cmn av).

41. Derived Adverb (drvd av); adverbs derived using the -ly suffix.

42. Compound Adverb (cmp av).

43. Numerical Adverb (numeric av).

44. Verbal Adverb (vbl av); verb appearing in an adverb position: He came running.

45. Verb Tag (tag v); adverbs of a prepositional type which accompany verbs: He called his girl up.

46. Quantifier-Intensifier (qnt-int).

47. Negative Adverb (neg av); the words not or no appearing in adverb positions.

48. Interjection Introducer (intj intro).

49. Expletive Introducer (expl intro).

50. Phrase Conjunction (phrs conj).

51. Clause Conjunction (cla conj).

52. Conditional-Resultive Conjunction (cnd-rsl conj).

53. Adverbial Conjunction (avbl conj).

54. Subordinate Conjunction (sub conj).

55. Conjunctive Pronoun (conj pn); also called relative pronoun.

56. Comparative Conjunction (comp conj); conjunctive words used in comparative structures: as big as ... or bigger than ....

57. Infinitive to (inf to); redundant with infinitive verbs.

58. Preposition (prep).

59. Infinitive Ambiguities (inf ambg); infinitives whose functions in the sentences are ambiguous.

60. Linking Ambiguities (aux ambg); forms of the verb be which are not definitely classifiable as either auxiliaries or linking verbs.

61. Participle Ambiguities (link ambg); participle verbs not clearly classifiable either as main verbs or verbal complements following a linking verb: He was concerned about.

62. Uncertain Classification (unc cls); idioms and completely ambiguous forms: They were flying planes.

Three features should be noted about this analysis. First, not all categories were of direct interest. Proper nouns, for example, were analyzed solely to remove their possible effects from the variables which were of interest. Second, the overlap between this analysis and the structure analysis may not always be as great as it would appear. The part of speech analysis was carried out using quasi-traditional methods which often produced results somewhat different from those produced by structural analysis. Third, all categories were mutually exclusive. Hence, the classification rules were ordered.

Anaphora Analysis: An anaphora is a pronoun-like structure in that it includes both a pro element and an antecedent. In fact, pronouns are one type of anaphora. Anaphora generally serve the function of allowing authors to state a complexly modified concept, set it equal to some shorter form, and thereafter refer to the complex concept using just that shortened form. In order for a reader to understand discourse, it seems necessary for him to have acquired some set of processes which enable him to identify anaphoric expressions of various types and correctly associate these anaphoric expressions with their proper antecedents.

Three kinds of variables based on anaphora seemed likely to correlate with the difficulty of prose. First, if the various types of anaphoric structures differ in frequency, they probably also differ in their familiarity to readers and therefore in the relative difficulty the reader has in inter[illegible] them. Hence, variables based upon the frequency with which a[illegible]ic structure appears in prose samples should correlate with th[illegible]ficulty of those prose samples. Second, the density with which anaphora occur should correlate with difficulty since each encounter with an anaphora involves some likelihood that an error of interpretation will occur. Third, as the time separating a reader's encounter with an anaphoric expression and its antecedent increases, the likelihood of his recalling the antecedent seems likely to diminish. This time interval can be roughly measured by counting the number of words intervening between an anaphora and its antecedent. This distance measure should correlate with the difficulty of prose samples.

However, the signs of the correlations between all anaphoric variables and prose difficultyare not easy to predict. It is true that anaphora probably confront the individual with additional comprehension operations to perform and that each additional operation probably entails an additional likelihood of error. However, the use of anaphora generally permits a reduction in the complexity of sentences. For example, Sentence (1) may be transformed into the two sentences The boy was small. He rode the horse. either of which is less complex than the original by most measures of complexity. Whether anaphora reduce more complexity than they introduce remains to be appealed to the data.

Since a detailed description of the anaphoric analyses used here is published elsewhere (Menzel, 1969), only illustrative examples will be provided here. In these examples, the anaphoric expression was placed in parentheses and its antecedent was italicized. Although an antecedent may consist of anything from a single word to a whole chapter, only short antecedents were shown.

1. Pro Anaphora (PRO AN); Bill left... (He) was..., Bill left. Joe (did) too.

2. Referential Repitition Anaphora (REF RPTN AN); Joe ate the cake. (Joe) liked it.

3. Formal Repitition Anaphora (FORM RPTN AN); The deer approached. The animal drank. Other (animals) followed.

4. Class Inclusive Anaphora (CLSS INCL AN); The horse grazed. This (animal) ...., Joe left. This (boy)....

5. Synonymous Anaphora (SYNM AN); Joe pushed the button. Then Bill (pressed) it.

6. Arithmetic Anaphora (ARTH AN); Joe and Bill left. The (former)....

7. Inclusive Anaphora (INCL AN); Bill beat Joe. (This) surprised....

8. Derivational Anaphora (DRVNL AN); Joe depends on Bill. Bill is (dependable))

9. Major Anaphora (MJR AN); Determined by selecting a nominal subject for the passage. Then, every anaphora of the subject is a major anaphora.

10. Minor Anaphora (MNR AN); The complement of the major anaphora.

Three types of variables were based on this analysis. The frequency variables were derived by counting the number of times an anaphora of a given type occurred in a passage and dividing it by the total number of anaphora. The abbreviations above were used to label these variables. An anaphora density variable, AN/W, was derived by dividing the total number of anaphora in a passage by the number of words in that passage. Finally, anaphora distance, AN DIST/W, was derived by counting the number of words intervening between each anaphoric expression and its antecedent, summing these numbers for all anaphora in the passage, and dividing by the number of anaphora in the passage.

Results

Analyses Performed: Three analyses were performed on the data. The first consisted of observing the sizes of the correlations between passage difficulty and each of the linguistic variables defined. The objectives of the analysis were to identify those variables which might cause passage difficulty and to identify those which might be used in practical readability prediction formulas. The second analysis consisted of examining the shape of the distribution of the scores on each variable. The object was to obtain the information necessary for selecting an appropriate model for calculating the readability formulas. In the past readability formulas have been calculated using methods which assumed that the relationships between linguistic variables and passage difficulty are best described by a straight line. This assumption has been shown to be erroneous (Bormuth, 1966). The shape of the regression curve which best describes the relationship between two variables is determined by the shapes of the score distributions of the two variables. For example, when the two distributions have identical shapes, the curve is a straight line. But if one distribution is more skewed than the other, the regression line is quadratic, having the shape of a more or less flattened c-shaped curve. Or, if one distribution exhibits more kurtosis, is more humped or flattened, than the other, the regression line will be cubic having the shape of a flattened s-shaped curve. The third analysis consisted of factor analyzing the linguistic variables, calculating a set of factor scores to represent each factor, and then correlating the factor scores with passage difficulty. The object of this analysis was to attempt to provide a simple description of the dimensions of prose which influence its difficulty.

Vocabulary Variables: Table 1 presents the Pearson product moment correlations between passage difficulty and each of the vocabulary

Insert Table 1 about here

variables as well as the statistics describing the distributions of these variables. All of these correlations were well above the levels necessary to be regarded as significant.

Many of the distributions differed significantly from the distribution of the passage difficulty scores. Skew and kurtosis were estimated by calculating the third and fourth moments of the distribution. The standard error of the skew estimates was .134 and the standard error of the kurtosis estimates was .268. These error estimates hold for all distributions reported throughout this section. The skew and kurtosis of the passage difficulty score distribution were .17 and -.62, respectively.

Structure Proportion Variables: Table 2 shows the same statistics for the structure proportions. However, in order to interpret these

---

Insert Table 2 about here

---

results it proved necessary to also show the proportion of passages in which each structure failed to occur at all. Obviously, it is not possible to examine the correlation between a passage's difficulty and a linguistic feature when that linguistic feature failed to occur.

Since several of the structures occurred either not at all or very rarely, nonsignificant correlations which involve such structures must be interpreted as being ambiguous. The correlations may have been nonsignificant because there is actually no correlation between a variable and passage difficulty, or they may have occurred because a correlation actually exists but is too slight to be statistically significant when so many zero scores occur. However, a significant correlation is interpretable as such regardless of the number of zero scores involved.

Hence it was necessary to interpret the results in Table 2 as forming three classes of variables: the variables having a status indicated by UC were interpreted as being uncorrelated with passage difficulty; the variables having a status indicated by the symbol C were interpreted as being correlated with difficulty; and the variables having a status indicated by a question mark were interpreted as possibly having a correlation with difficulty but one which was too slight to be detected from data where so many zero scores occurred on that variable.

The skew and kurtosis of many of the structure proportions differed significantly from the skew and kurtosis of the distribution of passage difficulty scores. In some cases, such as the tag question, these effects were primarily attributable to the rarity with which the structure occurred. But even in the cases of frequently occurring structures such as pronouns, the differences in skew and kurtosis were marked.

Syntactic Complexity Variables: Table 3 displays the statistics obtained from the analysis of the syntactic complexity variables. All

Insert Table 3 about here

of the variables except transformational operations per structure, and minimal punctuation units per structure correlated significantly with passage difficulty. And, again, the distributions of most of the variables differed from the distribution observed for passage difficulty.

Comparisons among the variables indicated that the highest correlations were obtained when the minimal punctuation unit was used as the syntactic unit analyzed. This was true regardless of whether the number of structures, transformational operations, or some other unit was used in the numerator. The outcome of the comparisons of correlations associated with each of the alternative methods of deriving Yngve depth was somewhat more complex. It appears that when the Nida phrase structure analysis is used, a count of left branches yields the highest correlations but that, when the modern phrase structure analysis is used, counts of right branches produced the highest correlations. Relative net depth measures seemed to yield the lowest correlations while depth per sentence seemed to yield the highest.

Part of Speech Variables: Table 4 shows the statistics obtained from the part of speech variables. Again, because some of the categories

---

Insert Table 4 about here

---

exhibited zero scores on many of the passages, it was necessary to distinguish three levels of status of the findings. Only articles do not correlate with difficulty. It may be that basic and designative adjectives also fail to correlate with difficulty because they did occur in nearly all of the passages. However, since it cannot be said with certainty that the correlation would have remained nonsignificant had these categories appeared in even those few passages, it is impossible to claim that these variables do not correlate with difficulty.

Anaphora Variables: Table 5 presents the results of analyzing the anaphora variables. All of the correlations were significant except

---

Insert Table 5 about here

---

those involving pro anaphora, arithmetic anaphora, and the major anaphora. Since no zero scores occurred for the major anaphora variable, it seems that no correlation exists between this variable and difficulty. Since minor anaphora constitute the complement of major anaphora it was unnecessary to examine the statistics of both variables. The distributions of most of the anaphora variables differed from the distribution of passage difficulty in both skew and kurtosis.

Factor Analysis: It was originally planned to factor analyze the matrix of correlations among the linguistic variables having significant correlations with passage difficulty, to calculate a set of factor scores corresponding to each factor, and then to correlate the factors with passage difficulty. For reasons explained in the discussion section which follows, factor scores were calculated for only 10 of the factors which emerged. Another change was that, in addition to the 94 variables which have already been defined and which met the criterion of correlating with passage difficulty, two other variables were included. The first was the ratio of lexical to structure words, WL/WS, and the second was the proportion of lexical words, WL/W. Since precise definitions will be given for these variables in another section it is presently sufficent to describe lexical words as consisting roughly of nouns, verbs, adjectives, and adverbs while structure words consist of pronouns, modal and auxiliary verbs, articles, prepositions, and the like.

In the first analysis a principal components model was used to extract all factors having eigen values greater than one. This resulted in 20 factors which accounted for 73.7 per cent of the variance. All scores were normalized before the calculations were performed. An orthagonal rotation was then performed on the factors using verimax procedures. The entire table is too unwieldy for publication, however, some of the more interesting results appear in Tables 6 and 7.

The results would have to be described as being quite complex. That is, instead of a few underlying factors accounting for nearly all of the variance, 20 factors had to be used to account for only 74 per cent of the variance. And, instead of all the variables exhibiting a high degree of communality, over one fourth, 25, had common variances of .6 or less, virtually assuring that many more common factors are required to adequately describe linguistic variables.

There were two patterns of factor loadings clearly discernable in the factor matrix. Nearly all of the syntactic complexity variables loaded heavily, .7 or above, on three factors while the vocabulary variables loaded heavily on a single factor. Thus, in this pattern, a very few factors described a great many different variables. In the other pattern, the remaining factors were characterized as representing principally one type of syntactic structure and one or more part of speech categories or anaphora which usually accompany that structure. For example, the existential there is very nearly always followed by a linking verb and the existential there is usually an expletive introducer part of speech, hence those three variables defined Factor VII shown in Table 7.

This second pattern of loadings suggested that very little common variance existed within the part of speech, syntactic structure, and anaphora variables when each is considered separately. To explore this possibility, 29 of the part of speech variables, 19 of the structure variables, and 8 of the anaphora variables were analyzed in separate sets using Jöreskog's (1967) maximum likelihood factor analysis model setting the probability of a solution's fit at .20. Twelve factors emerged from the analysis of the part of speech variables but these factors were primarily singletons. That is, one variable would exhibit a very high loading, .8 or .9, on the factor while none of the other variables exhibited loadings of over .2 or .3 on that factor. Furthermore, 13 of the 29 variables exhibited unique variances of .7 or higher. The analysis of the 19 syntactic structures gave an even more complex result. Only 4 factors emerged and 14 of the 19 variables exhibited unique variances of .7 or higher. The analysis of the anaphora produced similar results.

In contrast, nearly all of the syntactic complexity variables showed their major loadings on just three factors. Table 6 shows these three

---

Insert Table 6 about here

---

factors and the factor loadings. Of special interest was the fact that the Yngve depth measures based on counts of left branches seemed to be closely related to measures of syntactic length and other counts based on the number of units in sentences and minimal punctuation units. On the other hand, Yngve depth counts based on counts of right branches, and especially those counts which were freed in some manner from the effects of syntactic length, tended to load primarily on Factor II, a factor which they alone seemed to define. The remaining syntactic complexity variables appeared on a factor not shown in either Table 6 or Table 7. The variables defining this factor and their loadings were structures per word, .78; transformational operations per word, .75; referential repetition anaphora, -.63; and pronoun structures, .54.

To summarize the results from factor analyses, then, a simple structure does not seem to underly the variables correlating with passage difficulty. It was true that only five factors accounted for nearly all of the variation of the syntactic complexity and vocabulary variables. It was also true that a number of other factors emerged which exhibited high loadings of the part of speech, syntactic structure, and anaphora variables. But what was important to note was the fact that these additional factors resulted merely because there were overlapping categories in the three sets of variables. When each set was analyzed separately, the common variances all but disappeared. It was also important to note that many of the individual variables in these three classes of variables were not represented in the factors because they had very little correlation with any variable in the matrix. Hence, the factors obtained failed to represent a number of the variables which may be involved in the comprehension processes.

Factor Score Correlations: The preceding analyses made it only partially evident that a factor analytic approach was not sufficient to simplify explanations of the features which may cause passages to vary in difficulty. The possibility remained that some of the factors might not be correlated with passage difficulty. Thus, though many factors might underly the linguistic variables, it remained possible that only a few of the factors were required in order to account for variation in passage difficulty. Consequently, it seemed advisable to calculate a few of the correlations between factor scores and passage difficulty.

Factor scores were calculated for ten of the factors and then each set of scores was regressed on passage difficulty using a stepwise, polynomial, multiple regression procedure. It was necessary to use a polynomial regression since the regressions were not linear. This could be forecast from the fact that although the factor scores were normally distributed, the passage difficulty scores were both slightly skewed and platykurtic. Both the quadratic and cubic terms were significant in most of these regressions.

Table 7 shows the results of these analyses. All ten factors exhibited significant correlations with passage difficulty, but none

Insert Table 7 about here

was by itself sufficient to account for more than about 26 per cent of the passage difficulty. The correlations with difficulty can be regarded as essentially partial correlations from which the effects of the other factors are partialed out. This is true because an orthogonal rotation was performed on the factor matrix before the factor scores were calculated thereby assuring zero correlations among the factor scores.

These correlations have a special interest for they demonstrate that syntactic complexity is a concept quite independent of the particular structures a sentence contains, that the notion of syntactic complexity is a complex concept, and that all of these concepts correlate with passage difficulty. The correlations involving factors V through X seem to verify the suspicion that no simple explanation of passage difficulty variance is possible.

Discussion

The object of the first step in this analysis was to develop a large number of linguistic variables which may be involved in language comprehension and then to determine which of these correlated with passage difficulty. It was anticipated that a number of these variables would fail to correlate with difficulty and could thereby be excluded from further consideration in building the theory of language comprehension. The attempt failed for two reasons. First, almost every variable developed correlated significantly with difficulty. As a result, the problem of constructing a theory became more complex rather than simpler. Second, most of those variables which did fail to correlate with passage difficulty could not be definitely excluded because the structures upon which they were based failed to occur in large proportions of the passages, forcing an ambiguous interpretation. Either the variable did not actually correlate with passage difficulty; it correlated with difficulty but the correlation was too slight to be detected using only 330 passages; or it correlated but this fact could not be detected because the variable produced zero scores on so many passages.

This is not to say, however, that all of the new variables developed and shown to correlate with difficulty must be regarded as variables which may cause language difficulty. Neither the part of speech nor the syntactic length variables may be regarded as causes of difficulty since none of them may be directly manipulated. Rather, it is only syntactic structures and anaphora variables which may logically be claimed to cause difficulty. In order to vary the counts of parts of speech in a passage, it is necessary to derive alternate forms of the structures and anaphora. Similarly, syntactic length is dependent upon the transformations underlying the anaphora and syntactic structures. While the complexity of a sentence can be varied more or less independently of the particular structures the sentence contains, it is nevertheless true that those variations must be made by altering the syntactic and anaphora structures. So it is tempting to assign syntactic complexity the status of a dependent linguistic variable. Yet the fact was that syntactic structure variables showed very little correlation with the syntactic complexity factor in the factor analyses. Hence, the placement of a structure within a sentence must be regarded as, itself, a linguistic variable which is both conceptually and statistically independent of the structures.

The theory of comprehension will probably have to be more complex than even these results would make it seem. All of the vocabulary variables were dependent variables. Word frequency, for example, is based upon observations of the relative frequency with which different words occur in language responses of people. The Dale lists are similar in that they were based upon measures of children's abilities to select the correct synonyms of the words. Far from explaining language comprehension, these variables merely pose the problem of what causes words to vary in frequency and difficulty. That is, they themselves represent phenomena which must be explained by the theory of comprehension. Even the word length variables are dependent variables. The number of letters or syllables in a word is dependent upon the number of affixes out of which the word is formed. Specifically, there exist structures underlying a word which are similar to the syntactic structures presented here. It is these structures which will probably come to be regarded as causing words to vary in length as well as in difficulty. Undoubtedly, the abstractness of words also plays an important role in comprehension, but when this set of abstractness variables is defined, they must be defined for precise logical analyses rather than as people's ratings of how abstract they feel the words are.

Another important class of variables that has been almost totally ignored both in the past and in the present studies is the intersentence syntax or discourse organization variables. It is true that the anaphora variables provide a beginning on this analysis but it now seems only a crude beginning. For example, the category of pro anaphora lumps together the pros of nouns, verbs, adjectives, and other structures ignoring the fact that these structures are often quite different in complexity and frequency of occurrence.

With respect to the third objective of these correlation studies, probably the best conclusion that can be drawn is that the objective was based on the faulty premise that the language comprehension skills would be found to be basically simple. These data rather strongly argue for the contrary proposition. It seems that there are a great many different structures which serve as stimuli for the language comprehension processes and that each of these requires a different process in order to interpret it correctly.

Some of the specific results of the factor analyses deserve special attention. The results of the factor analysis of the syntactic structures and syntactic complexity variables proved to be especially interesting. First, the fact that the syntactic structures showed negligible loadings on the factors defined by the syntactic complexity variables while at the same time the factors which they, themselves, defined, correlated with difficulty suggests that a syntactic structure has an effect on comprehension which is quite independent of where that structure appears in the context of a sentence. And, conversely, the syntactic complexity of a sentence has an effect on comprehension quite independent of the particular structures the sentence happens to contain.

The syntactic complexity concept also seems to warrant further analysis. The fact that these variables divided themselves into three distinct factors plus the fact that each factor correlated with difficulty provides an important consideration for constructing a description of sentence processing. One speculation seems reasonable. Because Yngve depth based on a count of right branches was more or less independent of Yngve depth counts based on left branches and because both factors correlated with difficulty, it should be suspected that the comprehension processes involve both a memory of anticipated structures in a sentence and a memory of some representation of structures not yet completed as the sentence is read. It should not be too difficult to verify this suspicion experimentally.

The fact that variables based on minimal punctuation units exhibited higher correlations with difficulty than similar variables based on sentences may be partially explained from these data. Occurrences of clause conjunction structures seemed to be associated with passages which were somewhat easier to understand but when a structure was used which lengthened a clause, the difficulty of the passage was generally greater. However, this situation is far from being completely explained, for some structures which increase the length of a syntactic unit were associated with the easier passages. The existential there, the subordinate sentence where the time is the same, and verb complements are examples. Also, some structures which decrease the length of a clause were associated with the more difficulty passages. Whatever the final explanation may turn out to be on this matter, these studies make it seem virtually certain that the previous practice of attributing grammatical difficulty to sentence length is not only illogical but contrary to fact.

In the absence of a more detailed analysis of vocabulary, it is difficult to explain why counts of syntactic length yield higher correlations with difficulty when they are based on syllables than when they are based on any other unit. However, one speculation seems especially useful in planning future research. It may be the morpheme rather than the word which constitutes the basic unit of syntax. Hence, words such as unhappiness, which is composed of three meaningful units mo modifying each other, contain a syntax which must be considered when assessing the syntactic complexity of a sentence. If this speculation is true, basing Yngve depth counts on morphemes should result in higher correlations with difficulty than counts based on words. And, by analogy, to the analysis of sentence structures, it would be reasonable to expect different types of word structures to correlate with difficulty quite apart from the syntactic structure of the words in which they appear.

The results of the factor analyses cast grave doubts on whether it is presently possible for a readability formula to exhibit simultaneously economic practicality, face validity, and predictive accuracy. The factor analyses showed that for a formula to have face validity it must contain many variables, otherwise the formula would leave out linguistic features which seem likely to cause variations in language difficulty. But if all of these variables were included in a formula, the formula's practical value would be reduced because it would require much expense to perform both the linguistic analyses and the mathematical calculations. Still more serious, a formula containing many variables would almost certainly lack much accuracy. A readability formula is really just a multiple regression equation in which linguistic variables serve as the independent variables and language difficulty serves as the dependent variable. Adding variables to such equations also adds the error normally associated with the estimation of the beta coefficients. At some fairly early point, the error added by each new variable comes to exceed whatever predictive validity the variable may have added. Hence, adding enough variables to obtain a formula having convincing face validity will result in an uneconomical formula having low predictive validity. Obviously, some sort of compromise has to be reached in a way which is not yet entirely clear.

To summarize this discussion, these studies show that any theory which sets out to explain the language comprehension processes must be far more complex than originally supposed. Many different and independent structures correlated with difficulty and it seems highly likely that future research will lead to the discovery of still more such features. Syntactic complexity seems to produce an effect on comprehension which is independent of the effects produced by the particular syntactic structures a sentence contains. Further, the concept of syntactic complexity appears more complex than hitherto supposed. It now appears that measures of complexity should take into account the possibility that comprehension involves the memory of structures which are not yet completed at a given point in the sentence as well as the anticipation of structures begun but not yet completed. Clause length produces an effect also. The fact that higher correlations were observed when syntactic length was measured in units of syllables than when measured in letters, words, and so on, was taken as grounds for speculating that words themselves may exhibit a complexity very similar to that of a sentence. Finally, the results of the factor analysis showed that some compromise must be reached between the face validity, practical utility and predictive validity in designing readability formulas.

CRITERION CLOZE READABILITY SCORES

The object of the next studies reported was to determine what score a student should receive on a cloze readability test in order for its passage to be considered suitable for use in his instruction. To date, there have been just three studies in this sequence and, as yet, the results are both tentative and incomplete. But the problem of rationally selecting a criterion score has been solved in principle and the third of these studies demonstrates the form of that solution. Further, in order to calculate the readability formulas contained in this report, a criterion score of some sort had to be used and the criterion cloze scores of 35, 45, and 55 per cent were adopted. The studies described in this section provide the sole basis available for the selection of those scores as criteria.

A criterion level of performance has great general importance for the design of instructional materials. It is probably neither possible nor even desirable to design instruction on which performance is at or even near perfection. Any attempt to do so would ignore the fact that learning curves go asymptotic well before perfect performance is reached, thus assuring that efforts to reach perfect performance in instruction will almost certainly result in outrageous expense without actually attaining that objective. Furthermore, those efforts seem highly likely to result in repetitious drill or instruction so simplified as to be viewed by students as being boring and insipid. Thus, it seems reasonable to suspect that efforts to achieve perfect performance in instruction would probably be accompanied directly by rapidly mounting financial costs and rapidly deteriorating student attention to the instruction. Obviously, it will require more than a handful of studies to determine exactly what level of performance maximizes the desirable outcomes of instruction while minimizing the undesirable outcomes. The studies reported here more than anything else simply articulate the problem and provide a demonstration of the form of the solution to this problem.

Calculating a readability formulas forces the investigator to at least recognize that criterion levels of performance must be established. The readability formula is actually nothing more than a multiple regression prediction equation which permits its users to count certain features of the language in materials, insert these counts into the equation and, when the equation is solved, obtain a number which tells the user how much reading ability is required in order for a student to exhibit a desirable level of performance on those materials. Obviously, it would be possible to beg the question of what constitutes a desirable level of performance simply by calculating the formula to predict the mean cloze score that some group exhibited on the passage. But this would negate the purpose of the readability formula for educators who have to decide whether or not to use the materials with their students. It is not very useful to an educator to know that the formula predicts that some group of students in the Minneapolis suburbs would make a given mean score on cloze tests made from the passage. What is useful to the educator is an estimate of whether his students have sufficient reading comprehension ability to perform satisfactorily on those materials. Only this information is of great importance in helping educators decide how to get the best return from the half billion dollars they spend annually on instructional materials.

Once a criterion level of satisfactory performance is decided upon, it is a fairly simple matter to calculate useful readability formulas. Nearly all schools administer standardized tests of reading comprehension achievement in order to assess their students' comprehension ability. These levels of ability are commonly expressed in terms of grade placement scores. When a criterion score is adopted, a passage can be assigned a number which represents the level of comprehension ability a student must have in order to reach that criterion score on a cloze test made from the passage. This is done by administering both the standardized achievement test and the cloze readability test to the same students, calculating a regression equation for predicting the grade placement scores from the cloze scores, and then calculating the reading achievement grade placement score required to attain the cloze criterion score. When this grade placement number is subsequently used as the dependent variable for calculating readability formulas, the formulas estimate for their users the minimum grade placement score a student must have in order to exhibit the criterion level of performance on the materials he is analyzing.

It is widely claimed by authorities in the area of reading (Betts, 1946; Bond and Tinker, 1967; and Harris, 1962; for example) that materials are suitable for use in a student's unsupervised study if he can correctly answer at least 90 per cent of the comprehension questions asked him after he has studied the materials. Materials are regarded as suitable for use in the student's supervised instruction if he can answer at least 75 per cent of the questions. The study of materials more difficult than this are claimed (without supporting evidence) to result in frustrations which cause the child to learn negative attitudes toward his instruction and to acquire faulty reading habits.

Since comparable criterion scores had not been established for interpreting the scores on cloze readability tests and since it is essential to have a criterion score of some sort in order to scale passages used for calculating readability formulas, it at first seemed necessary to determine what cloze scores were comparable to the criterion scores accepted for tests of the traditional type. The first two studies summarized briefly here were directed at that objective. However, it subsequently became evident that it would be even more desirable to develop a criterion score which was based upon theoretical and empirical considerations. The third study represents the preliminary efforts in that direction. Since the first two studies have been reported elsewhere in detail (Bormuth, 1967b and 1968a) the descriptions given here will be brief.

Comparable Cloze and Multiple Choice Criterion Scores: In the first study (Bormuth, 1967b) 50 item cloze tests and 31 item multiple choice comprehension tests were made for each of nine passages. Each test was given to a group of 100 students enrolled in grades 4 and 5. The cloze tests were administered first and without letting the subjects inspect the intact passages. Three days later the subjects were given each passage to read and this reading was immediately followed by the subjects taking the multiple choice test made from the passage. A prior experiment had determined that subjects who had taken cloze tests over these passages did not exhibit means or standard deviations on the multiple choice tests that differed significantly from the means and standard deviations of the scores made by another group which took only the multiple choice test.

The multiple choice scores were corrected for guessing. The multiple choice and cloze scores were then pooled to form a single set of multiple choice scores and a single set of cloze scores. From these two sets of scores a regression prediction equation was obtained and this equation was, in turn, used to calculate the cloze percentage scores comparable to scores of 75 and 90 per cent on the multiple choice tests. The first row of Table 8 shows the results. At the time this study was

Insert Table 8 about here

conducted, it was thought that the cloze score found to be comparable to the 90 per cent criterion might be too low since the scores on the composite multiple choice test exhibited a moderate ceiling effect.

Comparable Cloze and Completion Test Scores: In the next study (Bormuth, 1968a), the object was to determine what cloze scores were comparable to the 75 and 90 per cent criteria as measured by scores on completion comprehension tests administered during an oral reading test. It is in the context of oral reading comprehension tests that these traditional criteria are probably most frequently used (Betts, 1946). This investigation used all four forms of the Gray Oral Reading Paragraphs. Each form consists of 13 paragraphs and each paragraph is at a different level of difficulty. A subject was given cloze readability tests over two of the paragraphs at each level of difficulty and then an orally administered comprehension test over each of the other two paragraphs immediately after he had read each paragraph orally. The paragraphs were systematically rotated to counter balance differences among the difficulties of passages at the same difficulty level. The subjects, 120 in all, were drawn in equal proportions from pupils enrolled in grades 4, 5, and 6.

The cloze scores comparable to the criterion scores were determined by finding the most difficult level upon which a subject was able to obtain first a score of 75 per cent and then a score of 90 per cent on the comprehension tests, noting the scores made on the cloze tests at these respective levels, and then averaging the two sets of cloze scores determined in this way. The second line in Table 8 shows that the results were fairly close to those obtained in the study using multiple choice tests. Since it seemed likely that the cloze score comparable to the 90 per cent criterion on the multiple choice test was depressed by a ceiling effect, the score obtained in this second study seems to be a better estimate of the cloze score comparable to the 90 per cent criterion.

Both of these studies share the problem that the population of traditional comprehension test items used were not rigorously defined and then sampled to compose the tests. Although this would be possible to accomplish now (Bormuth, 1969a) it was not possible at the time. As a result, the criterion scores might be expected to vary somewhat from one study to another just as the result of biases introduced by different test writers. However, in view of the fairly close agreement between the two rather dissimilar studies, one would guess that this variation might be small. As a result, in the readability studies described in the next section, readability formulas were designed using 45 and 55 per cent as criterion scores on the cloze tests, convenient approximations of the criterion scores found in these two studies.

Cloze Scores Associated with Maximum Information Gain

A search of the literature revealed that the 75 and 90 per cent criteria could be traced to recommendations made by E. L. Thorndike (1917) who seems to have acquired the criteria from teachers who, in turn, seem to have obtained them from oral tradition. The next study reported was a pilot project which demonstrated that it is possible to select criterion scores on a rational basis and it made a preliminary determination of what should probably be regarded as the minimum level of cloze performance which should be attained by a student before the materials from which the cloze readability test is made can be accepted as suitable for use in his instruction.

Procedure: In broad outline, this study involved these operations: First, students were formed into pairs who were matched according to reading comprehension ability. Second, one member of each pair was given a cloze readability test over a passage to determine the difficulty of the passage for his pair. Third, in order to determine the amount of information the pair gained as a result of reading the passage, the second member of the pair was given a multiple choice test over the passage as a guessing test. A week later he was given the passage to study and immediately thereafter given the same multiple choice test. The information the pair gained was obtained by subtracting this pair member's score on the first testing from his score on the second testing. Each pair's information gain score was then plotted against the pair's cloze difficulty score and a polynomial curve fit to the distribution in order to determine how the cloze difficulty of a passage was related to the amount of information the student gains from the passage.

A 52 item cloze readability test was used as the criterion for matching the subjects. The test was made from a 263 word passage taken from a text on elementary psychology by Kretch and Crutchfield (195). This test exhibited a corrected split-half reliability of .84 for the subjects used in this study. In order to match the subjects, they were first ranked according to the size of their score on this test and then, starting at the top of the distribution, successive pairs of students were selected -- one being randomly designated as member X and the other as member Y of the pair.

Two other passages, designated passages A and B, were drawn from the same source and a multiple choice comprehension test and cloze readability tests were made from each. Passage A contained 469 words and passage B, 398 words. Five forms of a cloze readability test were made from each passage. These different forms were subsequently randomly assigned to subjects in such a manner that each form occurred equally often in the study. This was done to counter balance any effects which might arise if the items in one form differed in difficulty from the average difficulty of the items in all five forms. The test over passage A exhibited a split-half reliability of .92 and the one over passage B, a reliability of .89 for the subjects in the study. The multiple choice test for passage A contained 34 items and the one for passage B, 39 items. Each question had four alternative responses. These tests underwent three editorial revisions, each time trying them out on small groups of subjects. The multiple choice test made from passage A exhibited a split-half reliability of .84 and the one made from passage B, one of .86 for the subjects in this study.

Originally, it was planned to use just junior college students in the study, but it eventually proved necessary to use a much broader range of students in order to obtain suitable numbers of subjects representing each level of difficulty. Of the students tested, 25 pairs came from grade 3, 23 pairs from grade 5, 15 pairs from grade 7, 28 pairs from grade 11, 24 pairs from junior college, and 15 pairs were graduate students. Because of absences, the data reported for passage A are based on 129 pairs and the data for passage B, on 125 pairs. At the first testing session, the matching test was administered. At the second session all of the X members of the pairs took the multiple choice test made from passage A and the cloze test over passage B while the Y members of the pairs were taking the multiple choice test made from passage B and the cloze test made from passage A. At this session and the earlier one, the subjects were told that this was a study of how well students could guess on these different kinds of tests. Roughly one and one-half weeks later, the X members were asked to read passage A and then re-take the multiple choice test over it while the Y members were doing the same on passage B. No time limits were imposed during the testing.

Analysis: The cloze readability score for each pair was obtained by finding the percentage of cloze items the subject answered correctly. The pair's information gain score was obtained by first correcting the multiple choice scores for guessing using the formula correct responses minus one-third of the incorrect responses and then subtracting the score made on the first administration of the test from the similarly corrected score made on the second administration of the test over that passage. The gain scores were expressed as gains in percentage scores.

The regressions of the information gain scores on the cloze readability scores were analyzed separately for each passage. To do so, stepwise polynomial regression analyses were performed. In both cases the first three powers of the gain scores accounted for significant amounts of the variance. These multiple correlations were .69 for passage A and .62 for passage B. When the two polynomial curves were plotted and superimposed on each other, they appeared quite similar. A test of their similarity was performed by fitting the curve calculated from the data on one passage to the data calculated for the other passage. In neither case did the curve originally calculated for a passage account for a significantly greater amount of the variance than the curve calculated for the other set of data.

As a result of this analysis the two sets of data were combined and a single, eighth degree polynomial regression fit to the combined data. The use of the higher degree polynomial permitted the curve to both show the general form of the relationship yet reflect much of the error fluctuation in the data. Thus it provided a visual means of assessing the stability of the relationship between the cloze readability scores and the information gain scores. A plot of this curve is shown in Figure 1. This figure shows that subjects who were able to answer less

---

Insert Figure 1 about here

---

than 25 per cent of the cloze readability items were able to gain little information from the passages. For those scoring above this point there was a sharp rise in the amount of information gained. This rise continued until the cloze readability scores reached roughly 35 to 40 per cent, at which point the curve leveled off.

The leveling off effect was not entirely attributable to a ceiling effect on the multiple choice test. There were only 12 scores above 90 per cent on the second administration of the multiple choice test. The leveling off seems instead to have resulted more from the fact that there was a positive correlation of .42 between the scores a subject made on the first and second administration of the multiple choice test. Thus, while students for whom the passage was quite easy made high scores on the second administration of the multiple choice test, their high scores were in part attributable to relevant information they possessed before they had read the passage. So they seem to have gained little more information than the students for whom the passage was somewhat more difficult.

Discussion: These data show that it is probably possible to establish rationally a criterion for judging whether a passage is suitable for a student. This is shown by the fact that two different sets of materials were used yet the curves which resulted were quite similar. This provides an indication that there may be a fairly fixed relationship between cloze readability and information gain. The data also show that this criterion can be tentatively placed at roughly 35 per cent on a cloze readability test. However, the latter claim must be heavily qualified.

The most important qualification arises from the fact that no account was taken of the influence of passage difficulty upon affective responses in the determination of this criterion. It is desirable, of course, to provide students with materials from which they can gain information, but it is even more desirable to provide them with materials which they will also study without any more duress than is ordinarily involved in instruction. From this standpoint, it seems likely that a criterion of 35 per cent is too low. A cloze readability score of this size is roughly comparable to a student being able to answer only 60 per cent of the items on a multiple choice comprehension test even after he has studied the passage. Clinical observations generally show that students who are forced to study materials this difficult voice strong objections and exhibit signs of frustration and inattention. As a result, materials which are at or near the criterion of 35 per cent should be regarded as the most difficult materials from which a student is likely to attain any positive benefit. But those materials should also be viewed as ones which are probably too difficult for anything but extraordinary uses and as ones from which the student may acquire a number of negatively valued behaviors.

The second reason for viewing the 35 per cent criterion with some mistrust is the possibility that this criterion might vary depending on the reading ability of the student, the difficulty of the passage, individual differences among students, or some set of interactions among these variables. The speculation that such interactions might exist is not entirely a result of conjecture. Coleman and Miller (1968) who varied passage difficulty found some indications that information gain may decrease on very easy passages as well as on very difficult passages. Because both passages in the present study were fairly difficult for the students and of roughly equal difficulties, this effect probably would not appear in these data. Similarly, Kammann (1966) found that the student's temperament and the passage's difficulty both produce an effect upon the student's rating of the interest of the poem. Hence, adopting just a single criterion score may over-simplify the situation.

Finally, a number of technical considerations should discourage placing very much confidence in the 35 per cent cloze readability criterion. Only two passages were selected for use in this study. Neither the number of passages nor the manner in which they were selected is adequate to permit the results of this study to be generalized to all passages. Also, the manner in which the multiple choice tests were made was inadequate to prevent the introduction of systematic bias. The questions in these tests were written to test a range of comprehension skills, but no prescribed rules were followed in deriving the questions so the questions could not be drawn from an enumerable population of items. As a result there is no way to be certain that the results would have been the same had other writers made the multiple choice tests. Finally, giving the student the multiple choice tests both before and after he had read the passage may have biased the results.

Selection of Criteria for Calculating Readability Formulas

While there are ample grounds for being suspicious of the 35 per cent criterion, there are also good grounds for using it. First, it is the only criterion presently in existence having any rational grounds whatever supporting its use. Second, many of the arguments upon which mistrust of the 35 per cent criterion is based apply equally to the 45 and 55 per cent criteria traditionally accepted. The only thing recommending the 45 and 55 per cent criteria is their widespread use in practice, but in view of the fact that the present state of the art of criterion selection is quite primitive, it seems inadvisable to overlook even so questionable an advantage as this one. Consequently, the readability studies reported in the next section have followed the practice of calculating four versions of each readability formula. Three were based on the 35, 45, and 55 per cent criteria and one was calculated in such a manner that the formula's user could select his own criterion. The results of subsequent research aimed at establishing sound criteria can then be incorporated into the latter formula.

READABILITY FORMULAS

Objective

The purpose of the studies reported in this section was to develop a set of readability formulas which are useful for estimating the suitability of materials for students. These formulas included ones for estimating not only the difficulties of passages but also the difficulties of individual words and sentences. Further, different formulas were designed especially for scientific use, machine automated analyses, and manual analyses by either skilled or unskilled personnel. The data for the calculation of these formulas were obtained from the 330 passages described above.

It seemed desirable to calculate readab'' y formulas for estimating the difficulty of individual sentences and words. In the past, formulas specifically designed for estimating the difficulties of passages have been used to estimate the difficulties of words and sentences within passages because no other kind of formula avaliable was based on sound research. The formulas used, however, probably led to systematically biased estimates. The central limits theorem assures that variables obtained by averaging across a passage will be more normally distributed than variables derived from smaller language units, since virtually all of the basic counts on language exhibit sharply skewed and leptokurtic distributions. Hence, formulas calculated from variables derived from whole passages are unlikely to conform to the shapes of the curves relating the linguistic variables to the difficulty of those units.

Formulas are also used for purposes ranging from providing controls over materials used in experiments to making rough estimates of the suitability of materials as they are being edited. In experimental settings the amount of labor and the level of skill required of the analyst is a secondary consideration. In the case of schools, the skills required of an analyst become crucial. And in many editing situations where massive amounts of materials must be analyzed, the labor requirements are elevated to the status of a primary consideration. Hence, it seemed necessary to design formulas which are as accurate as possible yet are adapted to the special requirements of the various users.

Grade Placement Scaling of the Passages

There were two different objectives for the passage scaling operations. The first was to assign three grade placement scores to each passage -- one corresponding to each of the cloze criterion scores, 35, 45, and 55 per cent. Each set of grade placement scores then served as a dependent variable for calculating alternate forms of the readability formulas for estimating passage difficulty. The second was to calculate and plot a general function, hereafter called the passage grade placement formula, which yields the grade placement of a passage when any selected cloze criterion score and the cloze mean estimated by one of the formulas presented here are substituted into the equation. Because the problem of what cloze criterion score is most desirable is very much an open question, it seemed necessary to calculate readability formulas which merely estimated the cloze mean observed in this study and then to provide a second formula with which to estimate a grade placement score when any cloze criterion whatever is chosen by the analyst.

Procedure: Grade placement numbers were assigned to each of the passages following this three step procedure. First, analyzing each passage separately, the students' cloze percentage scores on a passage were correlated with their reading achievement grade placement scores. All 285 students who took all five forms of the test made from the passage were pooled for this analysis in order to eliminate biases due to differences among the test forms. A few spot checks showed that many of these regressions were curvilinear so the regressions were performed using a stepwise polynomial regression model. In the second step the polynomial regression equation was used to calculate a series of predicted grade placement scores which corresponded to the cloze percentage scores starting with lowest and ending with the highest cloze percentage score observed on that test. The intermediate values were the integer multiples of 5 falling within this range, thus producing series such as 13, 15, 20, ..., 65, 70, and 74. While this calculation yielded many grade placement scores for a passage, it also yielded the grade placement scores corresponding to the 35, 45, and 55 per cent criterion scores. These calculations were repeated for each passage.

In the third step the passage grade placement formula was calculated. It expressed the grade placement of a passage as a function of its cloze mean and whatever cloze criterion it may seem appropriate to select. Consider that each passage has associated with it three sets of scores -- the cloze mean (M), the criterion scores (C) consisting of the highest and lowest cloze scores observed on the passage plus the scores at each multiple of 5 falling within that range, and a grade placement score (GP) found to correspond to each of those criterion scores. A stepwise multiple regression was performed on these scores, repeating the passage means as often as necessary to obtain a number corresponding to each of the criterion scores. The GP scores served as the dependent variable and M, M^2, M^3, C, C^2, C^3, and the powers of the cross products CM, $(CM)^2$, and $(CM)^3$ served as the independent variables. The data from all passages entered this analysis. All of the independent variables contributed a significant amount of variance to the regression analysis.

Results: Most of the regressions between the cloze and reading achievement scores were curvilinear. This is shown by the fact that in all but 27 of the passages either the quadratic, the cubic or both transformations of cloze scores accounted for enough of the regression variance to be considered significant at at least the .05 level and, therefore, entered the equations. Both the squares and cubes of the cloze scores entered 113 of the regression equations. Table 9 shows the range

---

Insert Table 9 about here

---

and distribution of the zero order correlations. These correlations should be considered high in view of the facts that the cloze tests contained only 18 to 24 items each and the regressions were, for the most part, curvilinear.

The three sets of grade placement scores seemed to provide stable estimates of passage difficulty. This is shown by their intercorrelations presented in Table 10. All of the correlations were above .9.

Insert Table 10 about here

The lowest correlations involved the grade placement scores corresponding to the cloze criterion score of 55 per cent. This seems likely to have occurred because, on many of the more difficult cloze tests, scores as high as 55 per cent tended to be scarce and in one case totally absent. As a result the regressions were probably less reliable in that region.

The passage grade placement formula provided a close fit to the data. The correlation between the passage grade placement scores calculated with this formula and the grade placement scores calculated directly from the cloze and achievement tests was .978, and the standard error of estimate was .61. The equation for calculating the passage grade placement scores is shown in Figure 2 along with plots of the formula. Each curve represents the relationship between grade placement

---

Insert Figure 2 about here

---

and the selected criterion score for passages having the cloze mean score appearing on that curve.

Discussion: These analyses provided the dependent variables for calculating readability equations which accept the cloze criterion scores of 35, 45, and 55 per cent as representing a satisfactory level of performance. There did not seem to be much error associated with making the transformations necessary to derive these grade placement scores for each passage since all of the passage grade placement scores exhibited high intercorrelations and high correlations with the cloze means. Nevertheless using these scaled scores and the passage grade placement equation does involve some error. Whether or not the error is sufficient to make any difference in practical situations remains to be seen. The graph of the equation is offered merely as a convenience to allow analysts who wish to select their own criterion to avoid solving the passage grade placement equation repeatedly.

Method of Calculating Readability Formulas

Stepwise multiple regression procedures (Draper and Smith, 1967) were used to calculate the formulas. When the stepwise procedure is used, the effect is to select from a set of linguistic variables that subset which results in the best estimation of difficulty. Because many of the linguistic variables were curvilinearly related to difficulty, a consistent practice was made of including in each initial set of linguistic variables the squares and cubes of the variables as well as their first powers. Checks of selected variables showed that no power higher than the cubic accounted for significant amounts of the variables' covariances with difficulty when the first three powers had been considered.

Passage Level Formulas

Four sets of formulas were calculated for estimating passage difficulty. The first set was calculated without placing any, other than statistical, restrictions upon what variables could enter the equations, and the remaining three sets were designed for use by analysts having varying degrees of technical skill, equipment, and other resources. Within each set, one formula was calculated for estimating the cloze means of passages and the remaining three formulas were calculated to estimate the grade placement difficulty scores obtained by scaling passages using the 35, 45, and 55 per cent criterion scores. The latter three formulas provided potential users with the option of adopting any one of the three criterion scores which presently have some basis for support or of selecting some other criterion of suitable performance. The user who wishes to adopt his own criterion can do so by using a formula which estimates cloze mean and then converting this number to a grade placement score by using either Figure 2 or the passage grade placement formula.

Because the formulas for calculating passage difficulty will probably receive the most extensive usage, it seemed essential to provide evidence on the extent of their validities as predictors of the difficulty of an independent sample of passages. This evidence was obtained using somewhat comparable data from an earlier study (Bormuth, 1966). Briefly, these data relate to 20 passages of 275 to 300 words each which were drawn from five subject matter areas in such a manner that they represent just passages having Dale-Chall readability values falling within the range from grades 4 through 8. The five forms of the cloze tests made from these passages were given to five matched groups, each consisting of 139 students enrolled in grades 4 through 8. The 1957 version of the California Reading Test was used to obtain the scaled grade placements of these passages. The norms of this earlier version are generally easier than the norms of the 1963 edition which was used to scale the 330 passages on which the formulas were calculated. In all other important respects the data for these 20 passages were collected and analyzed following the same procedures which were used in the present studies.

Unrestricted Formulas: Table 11 shows the four formulas which

Insert Table 11 about here

resulted when the matrix of intercorrelations among all linguistic variables having significant correlations with passage cloze means was analyzed. The only constraint placed on the variables entering these equations was that each had to account for a significant amount of the difficulty variance. This was determined by calculating the partial correlation between a linguistic variable and difficulty while holding constant the remaining variables in the equation and then performing an *F* test to determine if this partial regression accounted for a significant amount of variance. The .10 level of significance was selected in this analysis since the principal object was to include in the equations every variable having a reasonably strong independent relationship with the difficulties of passages and only secondarily to test the hypotheses that the partial correlations were not significant.

Table 12 shows the validity statistics for each of these formulas. The formulas exhibited correlations approaching .9 with the difficulties

Insert Table 12 about here

of the 330 passages on which the formulas were calculated. However, when cross validated on the data from the 20 passages, the validity coefficients dropped appreciably. These drops were primarily a statistical effect due to the complexities of the formulas. The calculation of each beta coefficient was accompanied by some error. This was the error associated with attempting to estimate the true value of a linguistic variable in an entire instructional work from the sample passage drawn from it. While the regression model used did not take this source of error into account when the formulas were calculated, the error could be expected to reduce the sizes of the correlations in cross validation studies. However, the decreases in the sizes of the validity correlations were somewhat less than would have been the case had the cross validation been performed on shorter passages. Each of the 20 passages was at least twice as long as each of the passages used to calculate the formulas with the result that they yielded correspondingly more reliable estimates of both the linguistic variables and the cloze difficulties of the works from which they were drawn.

All of the formulas seemed to be linearly related to the difficulties of the 330 passages on which they were calculated. The linearity of each of these regressions was determined by calculating the predicted difficulty of each passage using a formula and then calculating partial correlations between the observed difficulties of the passages and the second, third, and fourth powers of the predicted difficulties while holding constant the correlation between observed difficulties and the first power of the predicted difficulties. Each $\underline{F}$ ratio shown was the highest $\underline{F}$ observed for these partial correlations. When the same tests of linearity were applied in the cross validation analyses, significant curvilinearities were observed. It seems possible that this fact was also attributable to the error introduced by including a great many variables in these equations.

Short Form of the Unrestricted Formulas: Because of the unreliability introduced by the length of the unrestricted formulas, it seemed necessary to obtain forms of these formulas which contained fewer variables. The ten linguistic variables to be considered in the short forms were selected on the basis of their correlations with difficulty, the number of different unrestricted formulas in which they appeared, and their frequencies of occurrence in passages, where the latter consideration was relevant.

Table 13 shows the results of these analyses. It should be noted that, while the multiple correlations by which these formulas were cal-

Insert Table 13 about here

culated on the 330 passages were somewhat lower than those observed for the unrestricted formulas, the correlations in the cross validation analyses were not only higher than those obtained for the unrestricted formulas but also higher than the multiple correlations obtained in calculating the formulas, themselves. This effect was expected since the data for the 20 passages used in the validation analyses were more reliable than the data on the 330 passages from which the formulas were originally calculated. Hence, using even longer passages in the validity studies would probably have resulted in still higher validity correlations.

It should be noted that the formulas estimating cloze mean and grade placement scaled for a criterion of 35 were probably slightly curvilinearly related to the difficulties of the 20 passages. However, in an examination of the scatter plots of the expected versus the observed difficulties of these passages, the curvilinearities were so slight that they were not visually detectable. The mean biases given in Table 13 also show that the predicted grade placement scores were consistently lower than those observed on the 20 passages. This effect was attributable to the fact that the norms of the 1957 California Reading Test were easier than the 1963 norms for the test.

Manual Computation Formulas: Table 14 shows the formulas designed for use by unskilled analysts who have little or no convenient access

Insert Table 14 about here

to computers. The validity correlations were higher than the multiple correlations obtained when the formulas were calculated. The validity correlations of these formulas, while generally high, were below those observed for the short form of the unrestricted formulas. The predicted difficulties resulting from these formulas were linearly related to the difficulties observed for the 20 passages. The observed difficulties of the 20 passages were, again, consistently higher than those predicted by the formulas.

Machine Computation Formulas: The variables considered for use in the machine computation formulas were either those which a computer could calculate directly from a literal keypunching of a passage or which it could derive by scanning the text to see if it contained certain words. The latter variables were further restricted to include only those variables which would require the machine to store a dictionary which is sufficiently limited in size to permit the entire computation to be performed efficiently in a medium sized computer's core using blocked tape input. Within these constraints, an effort was then made to select those variables which would provide the greatest accuracy when included in a readability formula.

Table 15 shows the readability formulas which resulted. The multiple correlations which were obtained in the calculation of the formulas were fairly high and the validity correlations were equal to or higher than those obtained from the short forms of the unrestricted formulas. Further, the passage difficulties predicted by the formulas were linearly related to and consistently lower than the observed difficulties of the 20 passages.

Insert Table 15 about here

Sentence Level Formulas

In calculating the formulas at the sentence level of analysis, it was not possible to use all of the variables derived for the passages. Since most anaphora refer to relationships among sentences, they could not be regarded as sentence variables. The structure frequency variables had to be excluded because they occur with such low frequencies that they have little practical utility for determining the difficulty of a particular sentence. The part of speech frequency variables occur with perhaps more adequate frequencies but only if some part of speech categories are collapsed to form a smaller number of categories. This was accomplished using the combinations shown in Table 16. In forming

---

Insert Table 16 about here

---

the reduced categories, it was possible either to collapse into the same category just those parts of speech whose correlations with difficulty exhibited the same signs or to collapse the parts of speech into categories compatible with linguistic theory. The latter course was elected because it avoided the problems of ad hoc theories.

Sentence Formulas: Originally, it was planned that four formulas would be calculated at the sentence level of analysis -- an unrestricted, a short form of the unrestricted, and machine and manual computation formulas. As it turned out the short form of the unrestricted formula also met the requirements necessary for a machine computation formula. The results of these analyses are shown in Table 17. The formulas were

---

Insert Table 17 about here

---

moderately valid as shown by their multiple correlations. However, it was not possible to cross validate these formulas since the data from the earlier study were not easily recoverable in the appropriate form. Nor did it seem advisable to calculate formulas for directly estimating the grade placements of sentences. Since the scaling operations would require that each student be given a score on each sentence, those scores would be based on few responses making them very unreliable. The sentence scores predicted by the formulas were linearly related to the observed difficulties of the sentences from which the formulas were calculated.

Minimal Punctuation Unit Formulas: Because the minimal punctuation unit seemed to provide a somewhat better characterization of the syntactic unit, it was thought necessary to examine the advisability of deriving formulas to predict their difficulties, also. Part of the results are shown in Table 18 where it can be seen that the formulas exhibited

---

Insert Table 18 about here

---

slightly lower validities than those calculated for sentences. It was also found that the formulas were nearly identical to those obtained at the sentence level. The beta weights varied somewhat in the two sets of formulas and mean Yngve depth based on a count of right branches in Nida type syntactic analyses failed to appear in the unrestricted formula for minimal punctuation units. But this variable had barely exceeded the F level required to enter the sentence formula.

The redundancy of the sentence and minimal punctuation unit formulas was not unexpected. Most minimal punctuation units are also sentences. There were 2319 sentences in the data from the 330 passages and 2495 minimal punctuation units, indicating that at least 2143 or 92 per cent of the sentences were also minimal punctuation units. This fact leads to the expectation that the outcomes of the two analyses would be quite similar. Also, the somewhat lower validity of the minimal punctuation unit formulas was expected. An importantn distinction between the sentence and the minimal punctuation unit is the fact that the minimal punctuation unit contains fewer words which leads to the expectation that counts based on the minimal punctuation unit would be somewhat less reliable and that formulas based on those counts would be correspondingly less valid. Because of the redundancy and the lower validity of the minimal punctuation unit formulas, it seemed unnecessary to present them here.

Word Level Formulas

The design of formulas for predicting the difficulties of individual words faces the designer with the problem of designing two different types of formulas -- one for estimating the difficulties of words as those difficulties are influenced by the context of sentences and one for estimating the difficulties of words while ignoring the influences arising from the contexts in which they appear. In the former case it is necessary to take into account the syntactic context of the word, its syntactic function and its position in the structure of the sentence as well as the characteristics of the word itself. In the latter case, only features of the word itself can be taken into account. Both types of formulas were calculated.

In order to derive a variable to represent the function of a word in its context, it was necessary to collapse the part of speech categories into two classes, structural and lexical words. The categories combined to form these two classes are shown in Table 19.

Insert Table 19 about here

The word level formulas are shown in Table 20. These formulas were calculated by selecting five random words from a passage. This was

Insert Table 20 about here

necessary since no available computer program was able to handle accurately the extremely large numbers involved in these calculations. It can be seen that the formulas are moderately valid and linearly related to the word difficulties from which they were calculated.

Discussion

Passage Formulas: The passage level formulas reported here represent several considerable improvements over the formulas heretofore available. First, the formulas reported here are considerably more accurate. The Dale-Chall formula (Dale and Chall, 1948) which was probably the best of the earlier formulas exhibited a multiple correlation of slightly over .7 with the difficulties of the passages from which it was calculated. On the other hand, the machine computation formula for estimating grade placements scaled for a cloze criterion of 45 per cent exhibited a correlation of .83 and a cross validation correlation of .92. Hence, the machine computation formula reduces the uncertainty of estimating passage difficulty by roughly 15 to 35 per cent which represents an increase in precision of from 30 to 65 per cent over the best of the earlier formulas.

However, much improvement remains to be made in the validity of formulas. To begin with, the most accurate formula derived in the present study can account for only about 85 per cent of the observed variation in passage difficulty leaving the remaining 15 per cent as a challenge to future investigators. Furthermore, even if the present formulas had been perfectly accurate in predicting passage difficulty, they would still lack the type of validity necessary to make them completely fool proof. For example, the machine computation formula seems to assert that passages containing short words which all appear on the Dale List of 3000 Easy Words and which contain only short sentences not incorporating modal verbs will necessarily be easy to understand. Yet nearly any experienced writer can easily produce passages which fit all of these specifications yet which are extremely difficult to understand. As long as this is true, it will remain possible for readability formulas to yield misleading results. Readability formulas such as the unrestricted formulas appear to be more valid in this respect but they are so complicated to apply that they are nearly useless in practical operations. Hence the challenge to future investigators is to develop formulas which are not only precise and foolproof but also practical to use.

Second, the passage formulas presented here have received a test of their validity before being presented for use. The author knows of no instance where an earlier readability formula was subjected to cross validation in spite of the fact that there were strong reasons for believing that the earlier readability formulas could not be generalized to apply to all instructional materials. The best of the earlier formulas were usually based on some passages in a reading workbook series and those passages could not be said to be a representative sample of the passages found in instructional materials. The formulas reported here were, both, based on a more adequate sample of instructional materials and tested for their ability to predict the difficulties of different materials tested on different students.

However, the validity of the grade placement predictions made by these formulas remains to be fully tested. The differences between the grade placement scores predicted for the 20 passages and those actually observed was explainable by the fact that the different test norms were used in scaling the two different sets of passages. And the difference observed roughly agreed with the author's informal observations about the relative difficulties of the two sets of norms involved, but the fact remains that it has yet to be determined that the absolute sizes of the grade placement scores obtained from the formulas reported here will agree with those which would be obtained by scaling a new set of passages using the same achievement test norms. In order for these grade placement estimates to be of practical value, the grade placements predicted should agree with the grade placements observed.

A third improvement represented by the formulas reported here is the fact that the grade placement scores the formulas assign passages, when they are shown to be accurate, are more easily interpretable by practitioners. Suppose, for example, that the analysis of a passage results in its being assigned a grade placement score of 6.3 for a formula scaled to a criterion of 45 per cent. The precise interpretation of that score is that students having reading achievement grade placement scores of at least 6.3 on the California Reading Test, 1963 norms, would probably perform at or above the criterion level on a cloze test made from that passage. But stated in the simplest operational terms, that passage difficulty score means that the passage should probably not be used with students whose reading achievement scores on the California Reading Test, 1963 norms, are below 6.3.

It may seem unduely limiting to scale .il of the formulas in terms of a single reading achievement test when it is well known that tests differ considerably in the difficulty of their norms. That is not a difficult problem. In designing these studies two options were considered -- scaling the formulas using several different achievement tests and thus proliferating the number of formulas or of scaling all formulas using one test and then leaving it to subsequent research to determine how scores on the California Reading Test compare with scores on other tests of reading comprehension ability. The latter alternative was selected as being the best use of available resources, since the primary need was for basic research into readability.

But the important fact is that the difficulty scores the formulas assign to passages have a precise interpretation. This has not been so with earlier formulas. Some formulas assigned levels to passages based on nothing more substantial than some vaguely defined impressions of the formula's author (see Flesch, 1948 and 1950, for example). Others such as Lorge (1948) and Dale and Chall (1948) based their formulas on the norms given in the McCall Crabb Standardized Test Lessons , norms for which no research has ever been published. Furthermore, the authors of these formulas differed in the criterion scores they used to scale the passages without either justifying their choices of criteria or warning the formula users of the possible consequences of adopting the particular criterion score they chose to force on the users.

Sentence Level Formulas: Three cautionary notes seem especially relevant to the use of the formulas for predicting sentence difficulty.

The most important is that the formulas have not yet been tested for validity, and therefore should be used with great caution. The second is that they may not solve the problem of determining the difficulties of all individual sentences. The sentences upon which these formulas were calculated all appeared in the context of passages. While these formulas may also hold for sentences appearing in isolation such as in picture captions, directions, and so on, there are neither logical nor empirical grounds for claiming that they do. Until either or both types of evidence are brought forward, the formulas can be recommended for use only with sentences in context and the user should recognize that he is employing formulas which have not been validated on data independently gathered.

Finally, it may be tempting to use either the passage grade placement formula or Figure 2 for converting sentence readability scores into grade placement scores. This practice is wholly unjustified. To begin with, the passage grade placement formula was based on passage cloze scores not on sentence cloze scores, so any attempt to generalize the use of the passage grade placement formula in this way is to be avoided. There are some positive reasons for suggesting that this practice would yield misleading results. The most important is the fact that the distribution of sentence and passage difficulty differ. The mean, standard deviation, skew, and kurtosis of passage cloze means were .39, .11, .18, and -.62, respectively, while the same statistics for sentence cloze means were .43, .15, .25, and -.37. These differences virtually assure that, if a sentence grade placement formula had been calculated, it would assign a given cloze value a somewhat different grade placement value than would be assigned by the passage grade placement formula.

Probably the most appropriate use that can presently be made of the sentence formulas is for examining text where the chief objective is to determine if sentences are unduly difficult. That is, a text may appear as a whole to have an appropriate level of difficulty yet contain numerous sentences which are incomprehensible. The sentence readability formulas would be useful for detecting this fact, provided their other limitations are kept in mind.

Word Level Formulas: Exactly the same limitations that are relevant to the sentence readability formulas also apply to the word readability formulas. That is, they were not tested for validity, they can be applied only to words appearing in connected discourse, and they may not be given grade placement scale values. Like the sentence formulas, their chief use is in identifying the words in a passage which are likely to present unusual difficulty.

SUMMARY

Objective

The purpose of these studies was to obtain some of the information essential for improving the effectiveness with which students acquire the knowledge encoded in the written language in their instructional materials. The first objective was to perform what seemed to be the necessary first step in any systematic attempt to develop a theoretical account of the comprehension processes which is sufficiently detailed to be of use for designing instruction in comprehension. This was to analyze the linguistic features which seemed likely to serve as stimuli in the comprehension processes, to base variables upon these features, and then, through logical analysis and analysis of the correlations between these variables and the comprehension difficulties of passages, make a preliminary determination of what linguistic features may stand in a causal relationship to the comprehension difficulty of language. Because of limits on available resources, the primary focus was given to the analysis of sentence syntax.

The second objective was to apply the knowledge of the correlations between difficulty and the linguistic variables to the development of practical readability formulas which educators can use to help them select materials having a level of difficulty which is suitable for their students. It is possible to increase the amount students learn from materials either by adjusting the materials to the students' capabilities or by selecting just those materials which are at levels of difficulty suited for the students. Readability formulas aid educators in making the latter decisions. But, because they are based primarily on correlational research procedures, they are of little use for adjusting the difficulty levels of the materials.

In the course of developing the readability formulas it became evident that the traditional criteria for judging the suitability of the difficulty level of materials was based solely upon arbitrary choice.. And it was obvious that readability formulas based on those criteria would have a rather dubious validity. Therefore, a third objective developed which was to construct a decision theory and then employ this decision theory for selecting a level of performance which represented the level of comprehension difficulty at which the negatively valued outcomes from reading a passage are minimized while positively valued outcomes are maximized. While the solution of this problem was only partially achieved in these studies, the form of the solution was attained.

Analysis of Linguistic Variables

The analysis of the linguistic variables was carried out in five steps. First, a large number of features of language were identified as possibly serving as the stimuli for the comprehension process. These consisted of a set of vocabulary variables representing various features of words, a set representing the frequencies with which parts of speech occur in passages, a set representing the frequency with which syntactic structures analyzable as transformationally derived occur in passages, a set which characterizes the syntactic complexity of sentences according to various theories of sentence processing, and a set based upon the complexity and frequency with which various anaphoric structures occur in passages.

Second, 330 passages, each about 110 words in length, were drawn from instructional materials using a stratified random sampling procedure. The grid used required sampling from materials used at the primary, elementary, junior high, high school, and college levels and from ten subject matter areas. Each passage was drawn from a different work and written by a different author. Third, each of the linguistic variables was derived for each of the passages. Fourth, the comprehension difficulty of each passage was determined by making five forms of a cloze test over each passage, administering each form to different groups matched for reading ability, and then calculating the mean of the percentage scores made by the students. Roughly 2600 students enrolled in grades 4 through 12 were used in this testing operation. They were divided into 50 matched groups and each group took one test form over each of 33 passages. Fifth, each of the linguistic variables was correlated with the cloze difficulties of the passages and the variables exhibiting significant correlations selected for further analysis. These further analyses consisted of factor analyzing the intercorrelations among the linguistic variables, calculating factor scores, and then correlating the factor scores with difficulty.

It had been anticipated that these analyses would simplify the problem of constructing the theory of the comprehension of language. The failure to realize this expectation was almost spectacular. It was reasoned that linguistic features could be eliminated from consideration if they failed to exhibit a significant correlation on the grounds that a feature cannot cause difficulty if it does not at least correlate with difficulty. However, most of the variables defined correlated with difficulty, and the interpretation of many of those which did not correlate was made ambiguous by the fact that often the features upon which those variables were based failed to occur in a substantial number of passages. This made it impossible to determine whether the failure to exhibit a correlation was due to the fact that, in reality, no correlation existed or merely due to the fact that the number of occurrences was too small to permit the correlation to be detected.

Furthermore, it had been anticipated that the factor analysis would still further simplify matters by demonstrating that the large numbers of seemingly different variables really represented just a small number of factors and, perhaps, only a portion of those factors were required in order to explain comprehension difficulty. In fact, twenty factors emerged and still many variables remained which both exhibited little relationship to those factors and yet exhibited correlations with difficulty. Separate analyses of the variables based upon the parts of speech, syntactic structures, and anaphoric structures revealed that little common variation existed within these sets. Even the syntactic complexity variables partitioned themselves among about four differentt factors. And, finally, when the factor scores were calculated for some of these factors and then regressed on the difficulty scores, each exhibited a significant correlation with difficulty.

The most obvious conclusion to be drawn from these results was that any theory of language comprehension will probably have to be complex. There are a fairly large number of independently manipulable features of language and almost all may be in some way involved in the comprehension processes. However, this conclusion must be modified by the fact that not all linguistic variables can be regarded as standing in a causal relationship to comprehension regardless of the magnitude of their correlations. For example, neither the part of speech variables nor the sentence length variables can be logically analyzed as variables which cause variation in difficulty because neither type of variable is, itself, manipulable. Rather, to vary the proportions of parts of speech or to vary sentence length it is necessary to employ sets of operations which transform the structures in sentences into other types of structures. It is these sets of transformational operations which cause part of speech proportions and sentence length to vary and therefore it is these same transformational operations, and not their effects on parts of speech and sentence length, which logically must be regarded as the causes of variation in difficulty. By this reasoning only the sentence structures, anaphora, and syntactic complexity variables may be regarded as causes of difficulty because only they may be varied. None of the vocabulary variables employed in these studies can be assigned the status of causal variables because they cannot, themselves, be varied. For example, word length cannot be varied by arbitrarily deleting a syllable or a few letters. Rather, word length is changed by altering the morphemic structure of a sentence using some set of transformational operations and it is only to those operations that causality may be attributed.

These results also showed that it will be extremely difficult to construct readability formulas which simultaneously exhibit face validity, good predictive validity, and economy for a user. In order for a formula to exhibit face validity, it should take into account all of the variables which independently influence difficulty. Otherwise, an author could write passages which appear easy when the formula is applied but which are, in fact, made quite difficult because of variables not incorporated in the formula. In view of the present results, a formula must include a very large number of variables in order to exhibit face validity. But it is difficult to attain accuracy with such formulas because the error associated with estimating beta coefficients and errors in measurement of linguistic variables are cumulative leading the formula to have poor predictive validity. Furthermore, as more variables are added to increase the face validity of a formula, the application of that formula becomes increasingly expensive, requiring increasingly higher levels of skill by the analyst and increasingly greater computation skill and clerical expense. Consequently, it appears necessary to trade off face validity to attain predictive validity and practical utility.

Criterion Cloze Readability Scores

In order to calculate a readability formula which estimates whether a passage is suitable for a student, it is first necessary to have some notion about the level of performance a student should exhibit on a passage in order for that passage to be considered suitable for him. Readability performance criteria have existed for quite some time but they pertained to performance on traditional comprehension tests, so the initial studies in this sequence were directed at determining what scores on cloze readability tests were comparable to the criterion scores adopted for use with traditional comprehension tests. However, it became evident that the traditionally accepted criterion scores were based on nothing more substantial than oral tradition and arbitrary choice, so the third study in this sequence represents a preliminary attempt to apply a rational decision theory to the selection of a readability criterion score for cloze tests.

In the first two studies the effort was directed at determining what cloze readability test scores were comparable to the scores 75 and 90 per cent on ordinary comprehension tests. In the oral tradition of classroom practice, these are said to represent the minimum levels of performance necessary for materials to be used, respectively, intthe supervised and unsupervised instruction of students. These studies involved the use of various regression techniques to determine what cloze scores were comparable to these criteria. The results seemed to show that the cloze score of 44 per cent was comparable to the 75 per cent criterion and that the cloze scores of 57 per cent was comparable to the 90 per cent criterion. The cloze scores of 45 and 55 per cent were thereupon adopted as convenient approximations to these criterion scores.

For the third study it was asserted that a passage was not suitable for a student unless it permitted him to exhibit a maximum amount of information gain. Hence, this study undertook the investigation of the relationship of the information a student gains from reading a passage to the cloze difficulty of the passage for that student. Information gain was measured by subtracting the score a student made on a test over a passage before he studied the passage from the score he made on that test after he had studied the passage. The cloze difficulty of the passage was estimated for each student by pairing him with another student having nearly identical reading ability and then administering a cloze test over the passage to that student. An analysis of the regressions which resulted showed that, for students able to answer less than 25 per cent of the cloze items, very little information gain was apparent. For students scoring above the 25 per cent level there was a sharp increase in information gain until the cloze score of roughly 35 per cent was reached. Students scoring at or above this point all exhibited roughly similar amounts of gain. Hence, 35 per cent on a cloze readability test tentatively seemed to represent a criterion for deciding whether or not students are able to exhibit a maximum of information gain as a consequence of reading the passage.

However, this interpretation is offered only with many qualifications. First, clinical observations show that students find passages at that level frustratingly difficult and very much easier passages frustratingly insipid. Students will avoid the study of such passages whenever possible. Consequently, it seems essential to take into account the affective responses of students before settling on a particular criterion score. When this is done, it seems likely that the cloze criterion score adopted will be somewhat higher than 35 per cent. There are also a number of technical considerations which prevent placing very much confidence in this score -- only two passages were used, the test items may have been biased, and so on. However, in view of the fact that this is the only criterion score whatever having a rational basis, it seemed necessary to use it along with the 45 and 55 per cent criteria in calculating the readability formulas. But because of the tenuous nature of all of these criteria, it also seemed necessary to calculate general forms of the readability formulas which would permit the formula users to adopt better rationalized criterion scores as they are developed.

Readability Formulas

The purpose of these studies was to calculate formulas for estimating the suitability of passages and of sentences and words within passages. The formulas were designed to be as accurate as possible and yet adapted to the needs and resources of various users. Also the formulas were designed to permit users to adopt either the 35, 45, or 55 per cent cloze criteria or to use a generalized formula which would permit them to adopt any criterion score they chose. Finally, in order to make the results of the formulas easily interpretable to educators, the formulas were designed to estimate the minimum reading achievement grade placement score necessary for a student in order to attain a selected cloze criterion score on a cloze readability test made from the passage being analyzed.

The first step consisted of scaling the 330 passages giving each passage a set of grade placement scores corresponding to various cloze criterion scores. This was done for each passage separately by correlating the students' cloze percentage score for the passage with his grade placement score on a standardized reading achievement test and then using the regression equation which resulted to calculate the grade placement scores comparable to the lowest and highest cloze scores observed on that passage and comparable to each of the cloze scores that represented multiples of five falling within that range. These operations resulted in each passage being assigned the three grade placement difficulty scores corresponding to the 35, 45, and 55 per cent cloze criterion scores. They also provided the raw data for calculating the passage grade placement formula, an equation which permits a user to calculate the grade placement score of a passage by substituting in it a cloze criterion score of his choice and the cloze mean of the passage as estimated by one of the formulas developed in these studies.

Four sets of readability formulas were calculated for predicting passage difficulty. One set, the unrestricted formulas, included every variable having a significant partial correlation with difficulty when the other variables in the equation were partialed out. The second set, the short form of the unrestricted formulas, were calculated by considering just ten of the variables entering the unrestricted formulas. The third set, the manual computation, formulas were designed for use by unskilled clerical personnel with a minimum of training. The fourth set, the machine computation formulas, were designed to include just those linguistic variables which could be derived efficiently by a properly programmed computer. Four forms of these formulas were calculated, three to estimate grade placements scaled to the 35, 45, and 55 per cent cloze criteria and one to estimate the cloze mean which could then be used in the passage grade placement equation to estimate a passage's grade placement given any criterion score of the user's choice.

These formulas were cross validated using 20 passages, each of which was 250 to 300 words in length. The data on these passages were gathered in an earlier study and analyzed in a comparable fashion. Unfortunately, the norms of the reading achievement test differed from the norms used to scale the 330 passages upon which these readability formulas were based so it was not possible to check the accuracy with which the formulas predict grade placement scores. But it was possible to determine the validity of the formulas for ranking an independently drawn sample of passages.

The unrestricted formulas exhibited correlations ranging from .86 to .89 with the difficulties of the 330 passages, but only of .67 to .80 when cross validated on the 20 passages. The low predictive validity of these formulas was expected. Each formula incorporated 16 to 22 variables. The estimation of this many variables and the calculation of this many beta coefficients necessarily entailed a considerable amount of error. However, the validity coefficient was not as low as would have been the case had each of the 20 passages been of the same length as each of the 330 passages. The added length of the 20 passages provided them with greater reliability. The short forms of the unrestricted formulas exhibited correlations ranging from .83 to .87 with the difficulties of the 330 passages. However, their correlations with the difficulties of the 20 passages ranged from .88 to .93. The manual computation formulas exhibited correlations for the 330 passages ranging from .79 to .81 and cross validity correlations of .82 to .84. The machine computation formulas exhibited correlations of .81 to .83 with the difficulties of the 330 passages and from .92 to .93 with the difficulties of the 20 passages. These regressions were all linear except those involving the unrestricted formulas and two of those involving the short forms of the unrestricted formulas. The curvilinearities involving the short forms of the unrestricted formulas appeared slight.

Four formulas were calculated for estimating the cloze difficulties of sentences within passages -- an unrestricted formula, a short form of the unrestricted formula, a machine computation formula and a manual computation formula. The short form of the unrestricted turned out to be identical to the machine computation formula. These formulas were not cross validated because the necessary data were not easily recoverable from the earlier study. Nor was it possible to design the formulas to predict grade placement scores. The unrestricted formula exhibited a correlation of .66 with the sentences in the 330 passages, the machine computation formula exhibi+ed one of .65 and the manual computation formula exhibited a correl f .62. All of these formulas were linearly related to the difficu' , the sentences.

A similar set of formulas was calculated to estimate the cloze difficulties of minimal punctuation units. However, these formulas were almost totally redundant with those for predicting sentence difficulty, containing almost the identical set of variables and having very similar beta coefficients. The chief difference was that these formulas exhibited slightly lower validity coefficients. These results undoubtedly occurred because 92 per cent of the sentences were also minimal punctuation units, but were shorter and, therefore, less reliable than sentences. Because they added little to the sentence level formulas, those calculated for the minimal punctuation unit were not presented.

Two formulas were calculated for estimating the difficulties of the words within passages -- one estimates the difficulties of words as those difficulties are influenced by the surrounding context and the other ignores the effects of syntax. These two formulas exhibited correlations of .53 and .52, respectively, with the cloze difficulties of the words and both regressions were linear.

The passage level formulas represented a considerable improvement over the formulas heretofore available. The machine computation formulas provide increases of 30 to 65 per cent over the validities of the earlier formulas. And, unlike the earlier formulas, those reported here have been both cross validated and tested for linearity. Finally, they provide grade placement difficulty scores which are easily interpretable by users. However, it should be emphasized that the accuracy of the grade placement difficulty predictions has not yet been tested. The sentence and word level formulas should be used with caution. None has yet been cross validated. These formulas cannot be applied to words or sentences appearing in isolation. Finally, it seems unjustified to attempt to use the passage grade placement formula to convert sentence or word difficulties into grade placement scores. The results would almost certainly be misleading. The best use of those two sets of formulas is for examining text to determine if it contains words or sentences which are unduely difficult.

IMPLICATIONS

Perhaps, the major implication of these studies is to show that the objectives of future readability research must be greatly broadened. While the results are of immediate practical and theoretical interest and while they achieve the original objectives, they fall short of providing the ultimate solutions to the problems to which they are addressed. This almost had to be the outcome because at the time the main data collections were being planned and executed, the forms of the ultimate solutions and even the problems themselves appeared different than they appear at the time of this writing. This research led to a clearer understanding of the problems, causing the statements of the problems to be modified, and these reformulated problems, in turn, required somewhat different data for their solutions. Consequently, while these studies served very well to achieve their original objectives, they have also served to show how much broader these objectives should have been and how they might better be achieved.

Objectives of Readability Research

Traditionally, the approach to the study of readability has been both pragmatic and parochial. As it was conceived, the objective was to supply teachers with formulas which were simple enough for them to use without special training in linguistics and without complicated computation. The reasoning was that teachers were faced with an urgent practical problem and that researchers should attempt to solve that problem in a practical manner. Because of the teacher's limited background in linguistics, researchers have generally shown little inclination to gain an understanding of general language theory, presumably on the grounds that the linguistic variables to be derived from that body of theory would require more sophistication than could be expected of the teachers who would use the formulas. This seems to have been a realistic assessment of the teacher's situation and the same needs and conditions still prevail.

However, this traditional conceptualization of the problem can now be seen to be entirely too narrow. The educator's problem is actually to transmit knowledge to students using language, usually in written form, as the medium of communication. The effectiveness of this process can be improved either by improving the student's ability to comprehend language or by controlling the difficulty of the language in which knowledge is encoded. And controlling difficulty of language can be accomplished both by manipulating the language to make it more or less difficult and by selecting just those materials written in suitable language. Whereas traditional readability research was conceived as being relevant to just the selection of materials written in suitable language, modern researchers must now regard readability research as being vital to the solution of every major aspect of the problem of increasing the effectiveness with which students acquire the knowledge encoded in the language appearing in their instructional materials.

The efforts to analyze and describe the comprehension processes must succeed, at least partially, before it will be possible to design effective systematic instruction in those processes. But the development of the theory of language comprehension cannot proceed until researchers have identified the linguistic features which serve as the stimuli upon which the comprehension processes operate. For example, the theorist must know whether the grammatical structure of sentences is involved in the comprehension processes and whether the various features of that structure are involved differently. This is precisely the problem readability research is concerned with.

Readability research should also be concerned with the problem of manipulating language to make it suitable for students. It can hardly be regarded as sufficient just to provide educators with a mechanism which permits them only to select or reject materials on the basis of the suitability for students of the language the materials contain. To do so amounts to claiming that there is always an array of materials which differ in difficulty but otherwise provide the same instruction. Often, suitably understandable materials are simply unavailable, and when they are, they often provide instruction of a type that the educator does not prefer. Further, this ignores the enormous wastes associated with preparing materials which turn out not to be understandable simply because the publishers did not have an available technology for adjusting the materials to a suitable level.

In order to provide such a technology, readability research must be concerned not just with identifying variables which permit educators to predict the difficulty of materials but also with determining if those variables can be manipulated and, if so, if their manipulation produces effects upon the difficulty of materials. Specifically, readability research should be concerned with identifying the <u>manipulable</u> linguistic variables which stand in a causal relationship to difficulty.

Finally, it is questionable whether the best use of energy and talent is made when individual teachers are required to assess the difficulty of instructional materials. It would be definitely more desirable for private, state, or federal agencies to routinely determine and make public the difficulty levels of all instructional materials published. Such an agency could provide the expert analytic resources required to perform the detailed analyses which now seem essential. Furthermore, by centralizing this activity instead of having each analysis repeated by numerous individual educational institutions, the costs would be greatly reduced. It is probably true that teachers would still require simple, practical formulas for use with the special materials they prepare locally, but it has now been amply demonstrated that those simple formulas cannot fully meet the major needs of either the teachers or the publishers of instructional materials.

Thus, the objectives of readability research cannot be limited to just the short range attempts to provide recipes which are practical for individual teachers to apply. A formula which is practical for the individual teacher to apply probably does not give him an assessment of passage difficulty which has the face validity he actually needs. At most, such a formula provides him with a crude index of the information he really requires. But in order to provide the teacher with the information he needs, it will be necessary for readability researchers to recognize the fact that general language theory plays a central role in their research and then to use this theory for constructing readability formulas having satisfactory face validity. But these efforts of the readability researcher will come to nothing unless the various authorities in charge of allocating educational funds recognize that the classroom teacher has neither the time, training, nor resources to implement every technical advance made possible by research. The implementation of readability research depends heavily upon the special support given to agencies designed for that purpose.

Analysis of Linguistic Variables

The specific objectives of readability research, then, are to conceptualize the features of language which may serve as stimuli for the comprehension processes and subsequently to determine if and how these features affect the comprehension of instructional materials. Two matters should be discussed relative to this research. The first is to specify the relevance of the distinction between the manipulable and nonmanipulable linguistic variables. The second is to point out how this distinction affects the research methods used to study the two types of linguistic variables.

A feature of language is manipulable when there exists at least one other structure which is functionally interchangable with it and which regularly denotes essentially the same thing. For example, the words wide and broad in specifiable contexts and the passive and active forms of sentences such as Dogs chase cats. and Cats are chased by dogs. can be interchanged without appreciably altering the denotative meaning of the text. Similarly, the sentence structures presented in these studies and the sentences representing the base forms from which those structures were supposedly derived represent alternate forms for expressing approximately the same content. On the other hand, features such as sentence subjects and words denoting time do not seem to be manipulable because there are no known alternative forms in the language which permit a writer to express the same content but avoid using those forms.

The distinction between manipulable and non-manipulable variables is partially relative to the level of abstraction on which the analysis is carried out. The concepts of distance and transitive verb-agent-object sentences, for example, are probably non-manipulable features. Whereas, when they are analyzed at a less abstract level, some of their components such as wide and broad, high and tall and so on are manipulable. Every manipulable variable, then is a component of some non-manipulable variable. But it is probably not true that every non-manipulable variable can be analyzed into manipulable components.

It is important to conceptualize the non-manipulable features of language and to determine how those features affect the difficulty of instructional materials. Not only does this information make it possible to predict the difficulty students will have in comprehending materials, it also helps to identify those linguistic features on which a student should receive instruction and to determine the sequence of that instruction. For example, should student performance on language containing verb-agent-object sentences indicate that the presence of such sentences increases difficulty, it seems obvious that instruction should be devised which would increase the students' ability to interpret those features. Research is also required in order to determine the sequence of the instruction in the different skills so identified. Information about the relative difficulties of alternate forms of manipulable linguistic variables has all of these uses plus one more. Knowledge about their relative difficulties makes it possible to adjust the difficulty of instructional materials to make them understandable to students having varying amounts of comprehension skill.

Different research techniques are required depending on both the stage of the research and the type of feature being studied. It should be recognized that the conceptualization of linguistic variables is, itself, an important type of research. This type of research requires a considerable knowledge not only of structural linguistics but also of semantics, logic, and learning psychology, for the development of a linguistic variable is, in fact, the development of a psycholinguistic theory about how some portion of the comprehension process proceeds. Nor can this task be left to the linguist. Linguists have shown little concern for developing theories of linguistic complexity, for example. And, although linguistic theory does make use of the paraphrase alternations which constitute the manipulable variables of language, no effort is deliberately made to seek them out and catalogue them. Even the inventory of sentence structure alternations provided in these studies had to be devised specifically for these investigations even though the content with which they deal is of central concern to transformational-generative theory grammar, an area in which much linguistic research has been conducted.

When a variable, either manipulable or non-manipulable, has been devised, the next step should be to show that it correlates with comprehension difficulty. A variable cannot cause difficulty unless it at least correlates with difficulty, so it seems unjustified to risk the sizable amounts of funds and energy required to conduct an experiment unless it can at least be shown that the variable correlates with difficulty. Demonstrating such a correlation is becoming increasingly easy and economical. There is a growing number of language samples upon which difficulty data have already been gathered and published. However, there is presently a considerable need for a language sample in which the passages are of considerable length, say one or two thousand words each. This would greatly facilitate the economical examination of rarely occurring linguistic variables.

In order to establish the scientific basis necessary for altering the difficulties of materials the relative difficulties of alternate forms of manipulable variables must be contrasted directly through simple experiments. By definition, these alternate forms have the same structural function and approximately the same semantic meaning, so it is fairly clear to what source differences in difficulty should be attributed. However, this should not be construed as an assertion that alternate forms of manipulable variables have no differences in meaning. It is quite likely that all do differ somewhat with respect to connotative meaning and researchers should study these effects along with their study of how the alternations influence the comprehension difficulty of materials.

In order to sequence instruction on linguistic features and in order to predict passage difficulty, it is necessary to determine the hierarchical relationships among the features. An example of a hierarchical relationship is that existing between the nominalized verb structure and the derived anaphora as in <u>They are constructing a building. This construction is noisy.</u> where it appears to be necessary for a student to learn to interpret nominalized verbs of the type exemplified by <u>construct-construction</u> before he can learn to interpret this type of anaphoric expression. There are several research designs which make it possible to verify that these hierarchies do in fact exist (see Bormuth, 1969a). The simplest is an experiment in which one group is taught what is thought to be the simpler component skill, a second group is taught just the more complex skill, and then both groups are tested on both skills. When the skills are hierarchically related, teaching just the complex skill results in an increase in performance on both the simpler component skill and the complex skill relative to a control group taught neither skill. But the group taught just the simpler component skill improves only on the skill taught. This situation would produce a significant test-by-treatment interaction in a two factor analysis of variance. Instruction would ordinarily then be designed to proceed from the simple component skill to the complex skill.

It also seems necessary to determine the relative difficulties of the non-manipulable variables within each major class of variables even when they are not hierarchically related, since it is often desirable to sequence instruction beginning with the easiest structures and proceeding to the more difficult. This calls for special scaling procedures too complex for discussion here. However, they have been discussed in some detail elsewhere (Bormuth, 1969a).

Conclusion

The principal implication of these studies was that it is urgent to undertake the systematic analysis of the comprehension processes. Variables based on even the seemingly simple structures of language correlate with difficulty, strongly indicating that students have not mastered the language skills required to comprehend the content of language containing those structures. Readability research can do much to advance this analysis provided that investigators in this area shed the pragmatic and parochial approaches traditionally taken in this research. By making use of the analytic devices developed over the past three decades in linguistics, logic, and semantics, it now seems feasible to make major advances in the theory of language comprehension and those advances will make considerable improvement possible in the effectiveness with which language is employed as an instructional device.

REFERENCES

Ammon, P. R. The perception of grammatical relations in sentences: a methodological exploration. _Journal of Verbal Learning and Verbal Behavior_, 1968, _7_, 869-875.

Bar-Hillel, Y., Kasher, A., & Shamir, E. Measures of syntactic complexity. In A. D. Booth, _Machine translation_. Amsterdam: North-Holland Publishing, 1967.

Betts, E. A. _Foundations of reading instruction_. New York: American Book, 1946.

Bond, G. L., & Tinker, M. A. _Reading difficulties: their diagnosis and correction_. New York: Appleton-Century-Crofts, 1967.

Bormuth, J. R. _On the theory of achievement test items_. Chicago: University of Chicago Press, in press, 1969. (a)

Bormuth, J. R. An operational definition of comprehension instruction. In K. S. Goodman & J. F. Fleming (Eds.), _Psycholinguistics and the teaching of reading_. Newark: International Reading Association, 1969. (b)

Bormuth, J. R., Manning, J., Carr, J., & Pearson, D. Children's comprehension of between and within sentence syntactic structures. Paper read at American Educational Research Association, Los Angeles, 1969. (c)

Bormuth, J. R. Cloze test readability: criterion reference scores. _Journal of Educational Measurement_, 1968, _5_, 189-196. (a)

Bormuth, J. R. Effectiveness of current procedures for teaching reading comprehension. Paper read at National Countil of Teachers of English, Milwaukee, 1968. (b)

Bormuth, J. R. The implications of the use of cloze procedure in the evaluation of instructional programs. Occasional Report No. 3. Center for the Study of Evaluation of Instructional Programs, University of California, Los Angeles, 1967. (a)

Bormuth, J. R. Comparable cloze and multiple choice comprehension test scores. Journal of Reading, 1967, 10, 291-299. (b)

Bormuth, J. R. Readability: a new approach. Reading Research Quarterly, 1966, 1, 79-131.

Bormuth, J. R. Relationships between selected language variables and comprehension ability and difficulty. (United States Office of Education, Cooperative Research Project No. 2082) Los Angeles: University of California, 1964.

Chall, Jeanne S. Readability: an appraisal of research and application. Bureau of Educational Research Monographs, No. 34, Columbus: Ohio State University, 1958.

Coleman, E. B., & Miller, G. R. A measure of information gained during prose learning. Reading Research Quarterly, 1968, 3, 369-386.

Dale, E. A comparison of two word lists. Educational Research Bulletin, 1931, 10, 484-489.

Dale, E., & Chall, Jeanne S. A formula for predicting readability. Educational Research Bulletin, 1948, 27, 11-20.

Draper, N. R., & Smith, H. Applied regression analysis. New York: John Wiley and Sons, 1967.

Flesch, R. Measuring 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.

Fries, C. C. The structure of English. New York: Harcourt-Brace, 1952.

Gray, W. S., & Leary, B. A. What makes a book readable. Chicago: University of Cnicago Press, 1935.

Harris, A. J. Effective teaching of reading. New York: David MacKay, 1962.

Johnson, N. F. The influence of associations between elements of structured verbal responses. Journal of Verbal Learning and Verbal Behavior, 1966, 5, 369-374.

Jöreskog, K. G. Testing a simple structure hypothesis in factor analysis. Psychometrika, 1966, 31, 165-178.

Kamman, R. Verbal complexity and preference in poetry. Journal of Verbal Learning and Verbal Behavior, 1966, 5, 536-540.

Kretch, D., & Crutchfield, R. S. Elements of psychology. New York: Knopf, 1958.

Lorge, I. The Lorge and Flesch readability formulas: a correction. School and Society, 1948, 67, 141-142.

Lorge, I. Predicting reading difficulty of selections for children. Elementary English Review, 1939, 16, 229-233.

Martin, E., & Roberts, K. H. Grammatical factors in sentence retention. Journal of Verbal Learning and Verbal Behavior, 1966, 5, 211-218.

Mehler, J. Some effects of grammatical transformations on the recall of English sentences. Journal of Verbal Learning and Verbal Behavior, 1963, 2, 346-351.

Menzel, P. The linguistic bases of the theory of writing items for instruction stated in natural language. In J. R. Bormuth, On the theory of achievement test items. Chicago: University of Chicago Press, in press, 1969.

Miller, G. A., & Chomsky, N. Finitary models of language users. In R. D. Luce, R. R. Bush, & E. Galanter (Eds.), Handbook of mathematical psychology, vol. II, New York: Wiley, 1963.

Miller, G. A. Some psychological studies of grammar. American Psychologist, 1962, 17, 748-762.

Nida, E. A synopsis of English syntax. Norman, Oklahoma: Summer Institute of Linguistics, University of Oklahoma, 1960.

Rosenshine, B. New correlates of readability and listenability. Paper read at International Reading Association Convention, Boston, 1968.

Thorndike, E. L., & Lorge, I. The teacher's word book of 30,000 words. New York: Bureau of Publications, Teachers College, Columbia University, 1944.

Thorndike, E. L. Reading and reasoning: a study of mistakes in paragraph reading. Journal of Educational Psychology, 1917, 8, 323-332.

Yngve, V. H. A model and hypothesis for language structure. Proceedings of the American Philosophical Society, 1960, 104, 444-466.

Table 1

Correlations of Vocabulary Variables with Passage Difficulty
(N = 330)

Variable	r	Mean	S.D.	Differences	
				Skew	Kurtosis
LET/SYL	267*	3.085	.200	- .36*	1.43*
LET/W	-721*	4.482	.456	.27*	.70*
SYL/W	-683*	1.461	.195	.45*	.61*
TF/W	-509*	6.143	2.904	1.02*	2.00*
TFL/WL	-517*	10.031	5.090	.87*	2.32*
TFS/WS	-165*	1.355	.606	1.88*	5.03*
DSL/W	717*	.668	.114	- .29*	.16
DLL/W	727*	.819	.105	- .61*	.09

* An asterisk designates significance at the .05 level.

Table 2

Correlations of the Structure Proportion Variables with Passage Difficulties
(N = 330)

Variables	r	Zero Scores	Mean	S.D.	Differences Skew	Differences Kurtosis	Status
1 YES NO QUES	187*	95.1	.001	.003	8.91*	107.90*	C
2 WH QUES	112*	87.4	.002	.005	4.09*	23.77*	C
3 TAG QUES	-003	99.7	.000	.001	17.91*	325.62*	?
4 IMP SP	166*	90.8	.002	.007	6.67*	61.40*	C
5 IMP NSP	090	97.9	.000	.001	6.50*	43.27*	?
6 SEN NEG	081	59.8	.006	.009	2.07*	7.49*	?
7 CONST NEG	-231*	57.4	.007	.010	1.62*	3.77*	C
8 EXIST THERE	197*	73.3	.004	.008	2.12*	5.85*	C
9 CLEFT	-104	93.6	.001	.003	4.06*	19.63*	?
10 ANT IT	-014	89.0	.001	.003	2.73*	8.62*	?
11 PASSIV	-158*	35.3	.014	.016	1.87*	7.84*	C
12 CONST CONJ	-184*	8.0	.034	.024	.74*	1.42*	C
13 SEN CONJ	044	30.1	.012	.012	.93*	1.79*	?
14 PRENML N	054	32.5	.016	.018	1.11*	1.75*	?
15 PRENML AJ	-456*	5.8	.046	.033	1.24*	4.71*	C

Table 2 (continued)

Variables	r	Zero Scores	Mean	S.D.	Differences Skew	Kurtosis	Status
16 VBL AJ	-305*	54.6	.008	.012	1.74*	4.73*	C
17 NML AJ	024	89.3	.002	.007	5.06*	32.96*	?
18 RLTV CLA	-233*	12.0	.024	.017	.39*	.75*	C
19 PREP PH AJVL	-436*	6.8	.045	.029	.51*	1.21*	C
20 DRVD AV	-351*	46.3	.009	.012	1.57*	4.07*	C
21 PREP PH AVBL	-028	.3	.064	.027	.44*	2.03*	?
22 SUB SEN T-S	149*	59.5	.006	.008	1.42*	3.05*	C
23 SUB SEN T-B	040	92.0	.001	.004	6.91*	68.94*	?
24 SUB SEN T-A	063	95.7	.000	.002	5.06*	30.22*	?
25 SUB SEN IF	-002	80.7	.002	.005	2.18*	5.88*	?
26 SUB SEN TSH	069	91.7	.001	.003	3.34*	12.89*	?
27 SUB SEN CAUS	-081	78.5	.002	.005	1.76*	3.26*	?
28 SUB SEN PURP	111*	64.4	.005	.008	1.69*	4.16*	C
29 SUB SEN ALTH	-218*	88.7	.001	.004	3.16*	13.36*	C
30 COMP UNEQ	-006	66.6	.005	.010	2.86*	13.69*	?

Table 2 (continued)

Variables	r	Zero Scores	Mean	S.D.	Differences Skew	Kurtosis	Status
31 COMP EQ	-007	86.8	.002	.005	4.29*	26.12*	?
32 COMP SUPRL	-032	86.5	.002	.006	4.34*	27.68*	?
33 AJ CMPL	024	87.1	.002	.004	3.37*	16.07*	?
34 V CMPL	114*	42.9	.010	.013	1.80*	7.54*	C
35 N CMPL	-027	75.2	.003	.006	2.67*	12.88*	?
36 FACT NOM	132*	60.7	.007	.012	4.03*	26.98*	C
37 FOR-TO NOM	-006	86.5	.002	.006	3.73*	18.11*	?
38 POS-ING NOM	-074	74.9	.004	.009	2.96*	13.09*	?
39 CMPND N 1	-032	54.3	.008	.013	3.12*	7.55*	?
40 CMPND N 2	-142*	89.0	.002	.006	4.78*	32.83*	C
41 CMPND N 3	-017	42.6	.013	.018	2.18*	7.69*	?
42 CMPND N 4	-085	84.1	.002	.005	2.96*	11.31*	?
43 CMPND N 5	-006	95.7	.000	.002	6.25*	50.29*	?
44 CMPND N 6	017	96.0	.000	.003	7.07*	62.25*	?
45 CMPND N 7	063	88.7	.001	.005	4.01*	21.61*	?

Table 2 (continued)

Variables	r	Zero Scores	Mean	S.D.	Differences Skew	Kurtosis	Sta
46 PRONOUN	306*	10.7	.040	.034	1.24*	3.65*	C
47 DEL N	-051	63.5	.006	.010	2.37*	9.24*	?
48 S-A NEG	-	100.0	-	-	-	-	?
49 S-A QUES	-	100.0	-	-	-	-	?
50 RECIPROCAL	058	96.0	.000	.003	7.42*	71.82*	?

* An asterisk designates significance at the .05 level.

Table 3

Correlations of Syntactic Complexity Variables with Passage Difficulties
(N = 330)

				Differences	
Variable	r	Mean	S.D.	Skew	Kurtosis
Structural Density					
STR/W	-466*	.428	.066	- .18	1.00*
STR/CLA	-606*	3.906	1.464	.92*	2.03*
STR/MPU	-649*	7.142	3.469	1.00*	2.18*
STR/SEN	-640*	7.816	3.844	.97*	2.82*
Transformational Complexity					
TO/STR	-036	2.891	.292	5.14*	52.34*
TO/W	-423*	1.236	.220	.43*	2.48*
TO/CLA	-604*	11.273	4.273	.83*	1.60*
TO/MPU	-643*	20.652	10.261	1.04*	2.30*
TO/SEN	-640*	22.572	11.286	1.02*	3.40*
Strucutral Complexity					
STR C	-577*				
Yngve Depth					
NL/SEN	-497*	83.409	58.168	2.07*	9.34*
NR/SEN	-586*	31.823	17.932	1.02*	2.07*
ML/SEN	-526*	79.312	56.686	2.22*	11.22*
MR/SEN	-556*	35.088	20.452	1.27*	3.51*

Table 3 (continued)

Variable	r	Mean	S.D.	Differences Skew	Differences Kurtosis
NET(NL)	-452*	35.592	12.831	.91*	2.59*
NET(NR)	-491*	8.051	3.248	.72*	1.81*
NET(ML)	-526*	33.076	12.214	1.03*	3.47*
NET(MR)	-464*	9.781	3.748	1.18*	5.59*
NL/W	-453*	4.278	1.188	.85*	2.73*
NR/W	-516*	1.690	.315	.46*	1.29*
ML/W	-533*	4.039	1.118	.86*	3.28*
MR/W	-478*	1.853	.368	1.08*	5.73*
RND(NL)	-405*	.334	.111	.97*	3.23*
RND(NR)	-447*	.772	.287	.65	1.71*
RND(ML)	-497*	.317	.103	.92*	3.80*
RND(MR)	-409*	.317	.103	1.60*	8.81*
Syntactic Length					
LET/CLA	-626*	41.020	15.121	1.08*	2.31*
LET/MPU	-684*	74.609	34.523	.87*	1.47*
LET/SEN	-680*	81.549	38.295	.83*	1.67*
SYL/CLA	-627*	13.446	5.338	1.14*	2.61*

Table 3 (continued)

Variable	r	Mean	S.D.	Differences Skew	Differences Kurtosis
SYL/MPU	-689*	24.477	11.798	.83*	1.28*
SYL/SEN	-684*	26.766	13.054	.78*	1.33*
W/CLA	-527*	9.013	2.649	.86*	1.62*
W/MPU	-609*	16.341	6.682	.97*	2.21*
W/SEN	-605*	17.872	7.463	.83*	1.90*
CLA/MPU	-331*	1.833	.616	1.20*	2.85*
CLA/SEN	-325*	2.010	.731	1.28*	4.21*
MPU/SEN	-089	1.097	.166	1.90*	5.23*

* An asterisk designates a significance at the .05 level.

Table 4

Correlations of Part of Speech Variables with Passage Difficulty
(N = 330)

Variables	r	Zero Scores	Mean	S.D.	Differences Skew	Differences Kurtosis	Status
1 prp n	011	50.6	.018	.026	1.54*	3.30*	?
2 prp cmp n	-026	97.5	.000	.002	9.08*	104.00*	?
3 cmn n	-133*	0.0	.212	.046	- .24	.34	C
4 cmn cmp n	094	69.6	.005	.010	3.43*	18.33*	?
5 numeric n	138*	62.9	.009	.024	6.26*	49.40*	?
6 gerund	-111*	63.2	.006	.011	2.26*	6.02*	C
7 inf	148*	62.9	.005	.008	1.86*	5.56*	C
8 pers pn	512*	9.8	.037	.030	.76*	1.06*	C
9 dsgn pn	058	37.7	.012	.013	1.44*	4.39*	?
10 cmp pn	149*	78.5	.003	.006	2.40*	7.85*	C
11 ajvl pn	-021	96.9	.000	.002	7.46*	68.60*	?
12 vbl ajvl pn	-034	99.7	.000	.000	17.91*	325.62*	?
13 prp pos pn	050	99.7	.000	.001	17.91*	325.62*	?
14 cmn pos pn	-024	99.7	.000	.000	17.91*	325.62*	?
15 pers pos pn	033	98.8	.000	.002	12.42*	177.25*	?

Table 4 (continued)

Variables	r	Zero Scores	Mean	S.D.	Differences Skew	Differences Kurtosis	Status
16 cmp pos pn	-	100.0	-	-	-	-	?
17 trn act v	385*	2.5	.046	.029	.62*	1.02*	C
18 trn pas v	-268*	84.7	.002	.004	2.53*	9.01*	C
19 intrn v	248*	17.2	.022	.020	1.53*	5.71*	C
20 link v	241*	4.9	.029	.019	.70*	1.65*	C
21 cmp v	-029	98.5	.000	.001	7.79*	62.03*	?
22 aux v	-137*	47.2	.010	.012	1.10*	1.53*	C
23 modal v	252*	23.3	.017	.015	.94*	1.60*	C
24 pro v	078	94.8	.001	.002	4.76*	27.11*	?
25 cont v	117*	91.4	.001	.004	4.58*	26.59*	C
26 inf w-o to	184*	83.7	.002	.005	3.37*	18.93*	C
27 article	019	0.0	.106	.037	.23	1.05*	UC
28 dsgn aj	059	8.9	.026	.017	.32*	.28	?
29 bas aj	095	16.3	.027	.024	1.10*	2.92*	?
30 drvd aj	-630*	17.2	.030	.027	1.20*	3.28*	C
31 prp aj	-076	61.0	.010	.018	2.29*	7.79*	?
32 cmn aj	-138*	31.6	.018	.020	1.48*	4.01*	C
33 cmp aj	-092	84.7	.002	.005	3.00*	12.93*	?
34 numeric aj	055	49.4	.011	.016	1.87*	5.57*	?
35 vbl aj	-416*	19.9	.019	.017	.84*	1.35*	C

Table 4 (continued)

Variables	r	Zero Scores	Mean	S.D.	Differences Skew	Kurtosis	Status
36 prp pos aj	-021	92.9	.001	.003	4.05*	19.12*	?
37 cmn pos aj	-004	86.5	.002	.005	2.95*	10.90*	?
38 pnml pos aj	060	36.8	.013	.015	1.26*	2.80*	?
39 neg aj	066	85.9	.001	.004	2.47*	7.17*	?
40 cmn av	161*	22.7	.020	.019	1.09*	2.18*	C
41 drvd av	-381*	43.3	.010	.013	1.31*	2.98*	C
42 cmp av	144*	84.1	.002	.005	3.44*	17.26*	C
43 numeric av	048	92.6	.001	.003	3.70*	16.19*	?
44 vbl av	106	60.4	.005	.007	1.51*	4.40*	?
45 tag v	000	77.9	.003	.006	3.18*	18.02*	?
46 qnt-int	012	35.3	.012	.013	1.25*	2.83*	?
47 neg av	-015	62.0	.005	.008	2.27*	9.12*	?
48 intj intro	037	96.3	.000	.003	7.35*	47.54*	?
49 expl intro	137*	61.7	.006	.009	1.60*	3.39*	C
50 phrs conj	-189*	8.6	.029	.019	.38*	.57*	?
51 cla conj	-016	48.2	.007	.009	1.33*	3.29*	?
52 cnd-rsl conj	000	59.2	.005	.008	1.21*	1.93*	?
53 avbl conj	116*	26.7	.014	.013	.97*	1.92*	C
54 sub conj	-105	62.0	.005	.008	1.74*	4.70*	?
55 conj pn	-078	38.3	.009	.010	.77*	1.03*	?

Table 4 (continued)

Variables	r	Zero Scores	Mean	S.D.	Differences Skew	Kurtosis	Status
56 comp conj	-011	74.5	.003	.006	1.79*	4.47*	?
57 inf to	170*	30.1	.014	.014	1.12*	2.85*	C
58 prep	-432*	0.0	.115	.036	- .22	.56*	C
59 inf ambg	064	76.1	.003	.006	2.50*	9.16*	?
60 aux ambg	-024	35.9	.012	.013	1.67*	6.35*	?
61 link ambg	-029	34.7	.013	.015	1.76*	6.51*	?
62 unc cls	-059	77.6	.003	.008	2.45*	7.74*	?

* An asterisk designates significance at the .05 level.

Table 5

Correlations of Anaphora Variables with Passage Difficulty
(N = 330)

					Differences		
Variables	r	Zero Scores	Mean	S.D.	Skew	Kurtosis	Status
PRO AN	076	4.5	.249	.159	.71*	1.45*	?
REF RPTN AN	262*	3.9	.282	.153	.04	.60*	C
FORM RPTN AN	-116*	39.4	.062	.070	1.08*	2.19*	C
CLSS INCL AN	-130*	73.0	.022	.043	2.09*	5.36*	C
SYNM AN	-199*	76.1	.018	.039	2.38*	16.50*	C
ARTH AN	029	97.9	.001	.010	8.44*	80.27*	?
INCL AN	-110*	78.5	.012	.026	2.10*	5.52*	C
DRVNL AN	-265*	23.3	.110	.105	.98*	1.76*	C
MJR AN	-002	0.0	.224	.106	- .35*	.09	UC
AN/W	532*	-	.216	.086	.53*	2.16*	C
AN DIST/W	-392*	-	13.760	5.570	.95*	2.83*	C

* An asterisk designates significance at the .05 level.

Table 6

Loadings of Syntactic Complexity Variables on Three Factors

Variable	I	II	III
NL/SEN	94	17	-07
NL/W	93	-05	-09
ML/W	93	12	-01
NET(NL)	92	-05	-09
ML/SEN	92	24	-04
NET(ML)	91	12	-02
RND(ML)	91	08	-00
RND(NL)	89	-11	-08
W/MPU	85	31	-13
STR/MPU	81	29	-16
LET/MPU	80	29	-18
TO/MPU	79	28	-11
SYL/MPU	79	29	-19
STR C	77	13	-09
CLA/MPU	76	20	45
CLA/SEN	75	31	44
MR/SEN	75	58	-11
NR/SEN	74	59	-09

Table 6 (continued)

Variable	I	II	III
RND(NR)	29	84	-02
NR/W	39	82	-04
NET(NR)	36	82	-04
RND(MR)	33	78	-09
NET(MR)	41	77	-11
MR/W	43	77	-11
W/CLA	33	21	-75
LET/CLA	33	20	-71
RLTV CLA	35	02	70
SYL/CLA	33	20	-69
STR/CLA	38	21	-68
TO/CLA	38	20	-61
PRENML AJ	15	01	-45

Table 7

Correlations of Factor Scores with Passage Difficulty

Factor	Three variables having highest loadings on the factor		Polynomian R
I	NL/SEN	(94)[a]	
	NL/W	(93)	56
	ML/W	(93)	
II	RND(NR)	(84)	
	NR/W	(82)	16
	NET(NR)	(82)	
III	W/CLA	(-75)	
	LET/CLA	(-71)	20
	RLTV CLA	(70)	
IV	TF/W	(-85)	
	TFL/WL	(-83)	33
	DLL/W	(78)	
V	inf	(-85)	
	inf to	(-85)	13
	V CMPL	(-83)	
VI	drvd av	(85)	
	DRVD AV	(81)	32
	VBL AJ	(41)	

Table 7 (continued)

Factor	Three variables having highest loadings on the factor		Polynomial R
VII	expl intro	(89)	
	EXIST THERE	(88)	19
	link v	(39)	
VIII	YES NO QUES	(-60)	
	modal v	(-54)	19
	trn act v	(-53)	
IX	CONST CONJ	(86)	
	phrs conj	(81)	17
	aux v	(-37)	
X	cmn aj	(71)	
	DRVNL AN	(67)	16
	CMPND N 2	(58)	
XI	WL/WS	(-86)	
	WL/W	(-85)	44
	PASSIV	(-70)	

a. Numbers in parentheses show the loading of each variable on the factor indicated.

Table 8

Cloze Readability Test Scores Comparable to the Criterion Scores Used with Traditional Comprehension Tests

Type of Comprehension Test	Cloze Scores Comparable to		Standard Errors at	
	75 per cent criterion	90 per cent criterion	75 per cent criterion	90 per cent criterion
Multiple Choice	44.67	51.50	1.32	1.32
Oral Reading Completion Test	43.69	57.16	1.54	1.26

Table 9

Distribution of the Zero Order Correlations

between Cloze and Achievement Test Scores

(N = 285)

Size of Correlation*	Number Observed
80 - 89	7
70 - 79	133
60 - 69	136
50 - 59	45
40 - 49	7
30 - 39	2

* A correlation of .148 has a p .01.

Table 10

Intercorrelations Among the Cloze Means and Passage Grade Placement Scores

Variable	Mean	S.D.	Variable 1	Variable 2	Variable 3
1. Cloze Mean	.391	.114			
2. Criterion 35 GP	9.203	1.493	-.978		
3. Criterion 45 GP	10.446	1.492	-.978	.985	
4. Criterion 55 GP	11.537	1.446	-.937	.915	.963

Table 11

Unrestricted Formulas for Estimating Passage Readability

Independent Variable	Partial R	F^a	Beta
For Predicting Cloze Mean			
		Intercept =	.273921
$(DLL/W)^3$	413	63.40*	.205106
REF RPTN AN	318	34.68*	.138723
LET/MPU	-295	29.50*	-.002325
drvd aj	-260	22.47*	-.739366
$(LET/MPU)^2$	245	19.71*	.000009
cmn n	-237	18.39*	-.317875
AN/W	207	13.88*	.163294
EXIST THERE	206	13.75*	2.134094
$(numeric\ n)^3$	206	13.70*	13.731189
$(inf\ w\text{-}o\ to)^3$	185	11.00*	1,795.091553
trn pas v	-177	9.95*	-2.487564
cmn av	-163	8.42*	-.487778
$(expl\ intro)^3$	-160	8.08*	-1,296.053818
$(AN\ DIST/W)^3$	-150	7.10*	-.000001
RLTV CLA	-147	6.83*	-.484514
STR/W	-133	5.57*	-.133672
$(LET/SYL)^2$	-108	3.65*	-.020787
SUB SEN T-S	104	3.38*	.667857
$(LET/SYL)^2$	096	2.88*	.085992

Table 11 (continued)

Independent Variable	Partial R	F^a	Beta
For Predicting GP(35)			
		Intercept =	-92.000698
$(DLL/W)^3$	-383	52.77*	-2.607092
REF RPTN AN	-334	38.58*	-2.111183
LET/MPU	265	23.25*	.017633
EXIST THERE	-245	19.61*	-35.299893
drvd aj	226	16.50*	8.888002
AN/W	-217	15.14*	-2.398575
$(\text{inf w-o to})^3$	-215	14.88*	-29,002.042480
cmn n	194	12.02*	3.669046
$(LET/MPU)^3$	-192	11.73*	-.000001
trn pas v	183	10.60*	35.643973
numeric n	-178	10.09*	-6.521403
$(\text{expl intro})^3$	160	8.10*	17,934.875977
RLTV CLA	155	7.51*	7.015924
SUB SEN T-S	-144	6.46*	-12.848685
$(LET/SYL)^3$	134	5.62*	4.123199
cmn av	134	5.57*	5.543217
$(LET/SYL)^2$	-129	5.21*	-35.938873
LET/SYL	125	4.85*	104.339087
$(AN\ DIST/W)^2$	112	3.92*	.000480
$(CLS\ INCL\ AN)^2$	-095	2.77*	-10.667335

Table 11 (continued)

Independent Variable	Partial R	F^a	Beta
For Predicting GP(45)			
		Intercept =	-60.313025
$(DLL/W)^3$	-426	68.19*	-2.912971
REF RPTN AN	-412	62.84*	-2.382973
LET/MPU	273	24.89*	.029093
$(LET/MPU)^2$	-224	16.29*	-.000110
AN/W	-208	13.97*	-2.269171
$(numeric\ n)^2$	-204	13.31*	-39.094894
cmn n	202	13.15*	3.654701
EXIST THERE	-200	12.83*	-27.991251
$(inf\ w\text{-}o\ to)^3$	-198	12.63*	-26,261.886963
RLTV CLA	172	9.36*	7.726355
drvd aj	157	7.81*	12.588824
$(AN\ DIST/W)^2$	154	7.45*	.000656
cmn av	139	6.08*	5.698581
$(expl\ intro)^3$	132	5.47*	14,522.307129
trn pas v	118	4.34*	22.308641
SUB SEN T-S	-107	3.54*	-9.366940
$(LET/SYL)^3$	099	3.04*	2.992008
$(LET/SYL)^2$	-094	2.74*	-25.753284

Table 11 (continued)

Independent Variable	Partial R	F^a	Beta
For Predicting GP(55)			
		Intercept =	11.763236
REF RPTN AN	-334	39.19*	-1.945185
$(DLL/W)^3$	-323	36.34*	-2.946017
$(numeric\ n)^2$	-229	17.34*	-43.995816
$(AN\ DIST/W)^3$	221	16.04*	.000030
$(RLTV\ CLA)^3$	186	11.28*	2,348.849365
$(LET/SYL)^3$	183	10.85*	.030758
$(LET/MPU)^2$	-172	9.53*	-.000095
ML/SEN	166	8.89*	.006473
LET/MPU	147	6.95*	.019167
$(ML/SEN)^3$	-141	6.38*	.000001
$(inf\ w\text{-}o\ to)^3$	-137	6.02*	-19,096.886230
cmn n	135	5.78*	2.668448
$(DSL/W)^2$	-130	5.35*	-1.603100
AN/W	-129	5.26*	-1.419546
cmn av	115	4.22*	11.927396

a. All F values shown in this table were significant at the .10 level.

Table 12

Regression Analysis to Obtain the Unrestricted Formulas

Dependent Variable	Analysis (N = 330)				Validity (N = 20)	
	R	S.E.	F	Linearity F	r	Linearity F
Cloze Mean	.889	.052	64.95*	1.04	.738	8.88*
GP(35)	.888	.712	51.81*	1.17	.674	11.43*
GP(45)	.890	.702	56.14*	2.33	.721	10.06*
GP(55)	.861	.753	56.30*	1.97	.800	29.20*

* An asterisk designates significance at the .05 level.

Table 13

Short Form of the Unrestricted Formulas for Estimating Passage Readability

Score Predicted	Formula[a]
Cloze Mean	= .476824 + .254463(DLL/W)^3 - .003239(LET/MPU) + .000012(LET/MPU)^2 + .308649(REF RPTN AN) - .255777(REF RPTN AN)^2 + 15.906596(numeric n)^3 - .638471(drvd aj) - .200383(cmn n) - .002229(LET/SYL)^3 - .354143(RLTV CLA) + 4.583567(CLS INCL AN)^3
GP(35)[b]	= 8.624318 - 3.150610(DLL/W)^3 + .026493(LET/MPU) - .0000001(LET/MPU)^3 - 3.784290(REF RPTN AN) + 2.755072(REF RPTN AN)^2 - 29.198380(EXIST THERE) + 23607.204834(EXIST THERE)^3 - 8.234619(numeric n) + 2.382790(cmn n) + 7.829921(drvd aj) + .028063(LET/SYL)^3 - 14.824919(CLS INCL AN) + 5.605389(RLTV CLA)
GP(45)	= 9.293900 - 3.328815(DLL/W)^3 - 2.193862(REF RPTN AN) + .041174(LET/MPU) - .000152(LET/MPU)^2 - 2262.260315(EXIST THERE)^2 + 64687.930176(EXIST THERE)^3 + .027940(LET/SYL)^3 -214.446215(numeric n)^3 + 6.270336(drvd aj) + 2.698729(cmn n) + 6.012057(RLTV CLA) - 63.400957(CLS INCL AN)^3
GP(55)	= 10.865667 - 3.617262(DLL/W)^3 + .041881(LET/MPU) - .000156(LET/MPU)^2 - 1.684397(REF RPTN AN) - 48.532826(numeric n)^2 + 2170.182007(RLTV CLA)^3 + .022303(LET/SYL)^3 + 2.095581(cmn n)

Table 13 (continued)

Score Predicted	Formula Validity							
	Original Data (N = 330)				Cross Validation (N = 20)			
	R	S.E.	F	Linearity F	r	S.E.	Linearity F	Mean Bias
Cloze Mean	.874	.056	93.39*	.15	.913	.04	5.75*	--
GP(35)	.860	.778	68.89*	1.17	.911	.51	2.74	1.44*
GP(45)	.866	.760	79.07*	.00	.928	.54	4.16	1.68*
GP(55)	.833	.811	90.83*	.12	.877	.83	2.38	1.80*

a. The variables considered in these analyses were LET/SYL, DLL/W, LET/MPU, cmn n, numeric n, drvd aj, EXIST THERE, RLTV CLA, REF RPTN AN, and CLS INCL AN.

b. The program used to compute these equations rounded the coefficient of $(LET/MPU)^3$ to zero. When it was estimated using a different program, it was .000000083.

* An asterisk designates significance at the .05 level.

Table 14

Manual Computation Formulas for Estimating Passage Readability

Score Predicted	Formula[a]
Cloze Mean	$= 1.051674 - .099691(LET/W) - .004236(LET/MPU) + .000015(LET/MPU)^2$
GP(35)	$= .861207 + 1.279050(LET/W) + .050548(LET/MPU) - .000172(LET/MPU)^2$
GP(45)	$= 1.849494 + 1.307968(LET/W) + .053930(LET/MPU) - .000191(LET/MPU)^2$
GP(55)	$= -1.231834 + 2.764035(LET/W) - .023845(LET/W)^3 + .051591(LET/MPU) - .000186(LET/MPU)^2$

Formula Validity

Score Predicted	Original Data (N = 330)				Cross Validation (N = 20)			
	R	S.E.	F	Linearity F	r	S.E.	Linearity F	Mean Bias
Cloze Mean	.808	.067	204.91*	2.45	.833	.052	.57	--
GP(35)	.786	.927	175.45*	.00	.829	.79	.06	1.41*
GP(45)	.800	.899	193.13*	1.88	.838	.79	.01	1.71*
GP(55)	.785	.902	130.03*	.34	.823	.73	.05	1.99*

a. The variables considered in these analyses were LET/MPU, LET/W, pers pn, and prep.

* An asterisk designates significance at the .05 level.

Table 15

Machine Computation Formulas for Estimating Passage Readability

Score Predicted	Formula[a]
Cloze Mean	$= .886593 - .083640(LET/W) + .161911(DLL/W)^3 - .021401(W/SEN) + .000577(W/SEN)^2 - .000005(W/SEN)^3$
GP(35)	$= 3.761864 + 1.053153(LET/W) - 2.138595(DLL/W)^3 + .152832(W/SEN) - .002077(W/SEN)^2$
GP(45)	$= 3.398999 + 1.107014(LET/W) + .155327(W/SEN) - .002184(W/SEN)^2 + 6.672669(DLL/W)^2 - 7.523689(DLL/W)^3 - 5.266225(\text{Modal v})$
GP(55)	$= 3.450806 + 1.094841(LET/W) + .153830(W/SEN) - .002242(W/SEN)^2 + 11.478313(DLL/W)^2 - 11.224816(DLL/W)^3 - 5.427013(\text{modal v})$

Formula Validity

Score Predicted	Original Data (N = 330) R	S.E.	F	Linearity F	Cross Validation (N = 20) r	S.E.	Linearity F	Mean Bias
Cloze Mean	.833	.063	147.36*	.18	.920	.04	.39	--
GP(35)	.808	.884	153.13*	1.75	.916	.57	.29	1.49*
GP(45)	.825	.852	114.40*	.00	.927	.54	.72	1.70*
GP(55)	.808	.860	101.22*	.00	.915	.52	.77	1.91*

a. The variables considered in these analyses were DLL/W, LET/W, W/SEN, pers pn, prep, cmp pn, modal v, and inf to.

* An asterisk designates significance at the .05 level.

Table 16

Reduced Part of Speech Categories

Reduced Category	Contained Classes
Noun, (R n)	prp n, prp cmp n, cmn n, cmn cmp n, gerund, pers pn, dsgn pn, cmp pn, prp pos pn, cmn pos pn, pers pos pn, cmp pos pn, expl intro
Nominal Compound, (R nmnl cmp)	prp aj, cmn aj
Transitive Verb, (R trn v)	trn act v, trn pas v, cmp v
Intransitive Verb, (R intrn v)	intrn v, link v
Participle and Infinitive, (R prt inf)	inf, vbl ajvl pn, inf w-o to, vbl aj, vbl av, inf ambg, link ambg
Adjective, (R adj)	ajvl pn, bas aj, drvd aj, cmp aj, prp pos aj, cmn pos aj, pnml pos aj, neg aj
Adverb, (R av)	cmn av, drvd av, cmp av, numeric av, neg av
Article, (R art)	article
Pre and Post Article, (R p p art)	numeric n, dsgn aj, numeric aj
Qualifier, (R qlf)	qnt-int
Modal (R modal)	modal v, pro v
Aspect (R aspect)	aux v, cont v, aux ambg

Table 16 (continued)

Reduced Category	Contained Classes
Preposition and Participle, (R prep prt)	tag v, inf to, prep
Conjunction, (R conj)	phrs conj, cla conj, cnd-rsl conj, avbl conj, sub conj, conj pn, cmp conj
Misc, (R misc)	intj intro, unc cls

Table 17

Formulas for Predicting the Cloze Difficulties of Sentences

Score Predicted	Formulas

Unrestricted

$$\text{Cloze Mean} = .758835 - .176440(\text{R trn v}) + .163373(\text{R nmnl cmp}) + .126989(\text{R n})^2 - .078377(\text{R prep prt}) - .088749(\text{R av}) + .140650(\text{DLL/W})^2 + .065264(\text{DSL/W})^2 + .001319(\text{TF/W}) - .000976(\text{LET}) - .110237(\text{LET/W}) + .005888(\text{LET/W})^2 - .002265(\text{TQ/MPU}) + .000018(\text{TQ/MPU})^2 + .000023(\text{SYL})^2 - .011469(\text{NR/W})$$

Machine Computation and Unrestricted, Short Form

$$\text{Cloze Mean} = .703885 - .002675(\text{LET}) + .000011(\text{LET})^2 - .0000003(\text{LET})^3 - .054572(\text{LET/W}) + .099210(\text{DLL/W})^3 + .042741(\text{DSL/W})^2$$

Manual Computation

$$\text{Cloze Mean} = .578543 - .014349(\text{LET}) + .000054(\text{LET})^2 - .0000004(\text{LET})^3 - .047842(\text{W}) - .000748(\text{W})^2 + .000005(\text{W})^3$$

Formula	R^a	F	S.E.	Linearity F
Unrestricted	.663	120.34*	.115	1.02
Short Form of Unrestricted and Machine	.645	273.98*	.118	.39
Manual	.619	239.18*	.121	1.23

a. N = 2319.

* An asterisk designates significance at the .05 level.

Table 18

Regression Analyses to Obtain the Formulas for Predicting the Cloze Difficulties of Minimal Punctuation Units

Formula	R[a]	F	S.E.
Unrestricted	.645	116.46*	.118
Short Form of Unrestricted and Machine Computation	.622	174.55*	.121
Manual	.596	228.96*	.124

a. N = 2495

* An asterisk designates significance at .05 level.

Table 19

Definitions of Structural and Lexical Categories

Category	Contained Classes
Structural	link v, aux v, modal v, article, dsgn aj, neg aj, tag v, qnt-int, neg av, intj intro, expl intro, phrs conj, cla conj, cnd-rsl conj, avbl conj, sub conj, conj pn, cmp conj, inf to, prep, aux ambg, link ambg, unc cls
Lexical	all other categories

Table 20

Formulas for Estimating the Cloze Difficulties of Words

Score Predicted	Formulas
Words in Context	
	Cloze Mean = .900063 - .092991(LET) + .000382$(LET)^3$ - .063139(NR) - .006637(ML) + .009844$(NR)^2$ - .044190(L-S) - .035738(SYL)
Individual Words	
	Cloze Mean = .812807 - .100400(LET) + .000410$(LET)^3$ - .037406(SYL)

Formula	R^a	F	S.E.	Linearity F
Words in Context	532	92.77*	.263	.86
Isolated Words	522	205.50*	.264	1.17

a. N = 1646.

* An asterisk designates significance at the .05 level.

FIGURE 1

EIGHT DEGREE POLYNOMIAL CURVE FITTED TO THE REGRESSION OF EACH PAIR'S INFORMATION GAIN SCORE ON ITS CLOZE READABILITY SCORE

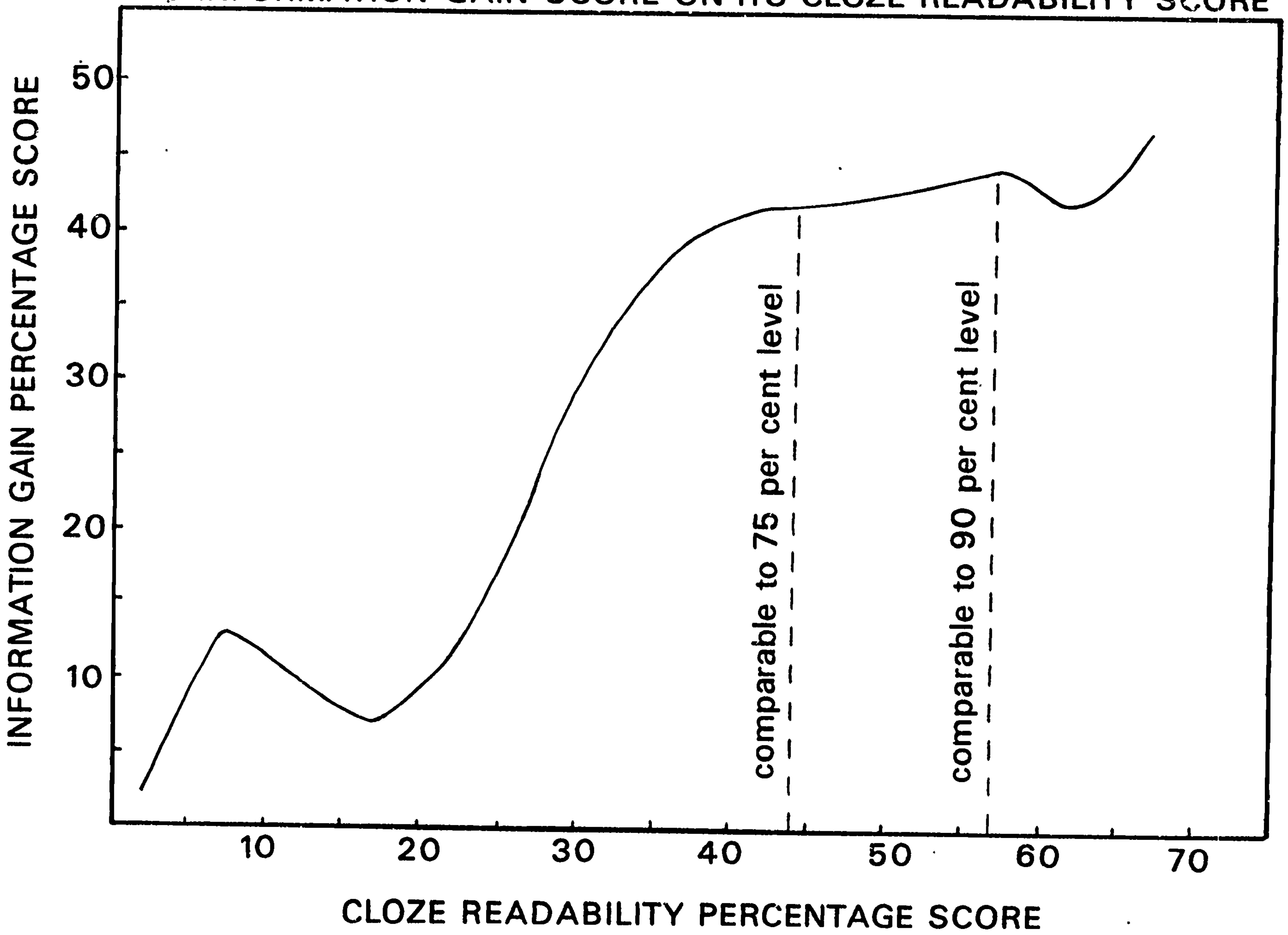


FIGURE 2

Graph for Converting Predicted Cloze Readability Scores into Grade Placement Difficulty Scores